MIXED MODELS: THEORY and COMPUTER PACKAGE OUTPUT*

by

S. R. Searle

Biometrics Unit, Cornell University, Ithaca, New York

Abstract

Theory associated with mixed models and estimation of variance components is summarized in note form. Output from computer packages SAS (routines VARCOMP, RANDOM and HARVEY) and BMD (P3V and P8V) is illustrated with sample pages from Annotated Computer Output (ACO) that is now available for these routines.

* Invited paper prepared for a Symposium entitled "Statistical Packages for Mixed Models with Unequal Subclass Numbers" at the Annual Meeting of the American Society of Animal Science, Raleigh, North Carolina, July 29, 1981.

Paper No. BU-741-M in the Biometrics Unit.

FIXED EFFECTS MODELS

Model: $\underset{\sim}{y} = \underset{\sim}{X}\underset{\sim}{b} + \underset{\sim}{e}$

$\text{var}(\underset{\sim}{e}) = R$ (usually $\sigma^2 \underset{\sim}{I}$)

fixed effects: unknowable constants, of particular interest, no sampling involved.

Estimation:

OLS: $\underset{\sim}{X}'\underset{\sim}{X}\hat{\underset{\sim}{b}} = \underset{\sim}{X}'\underset{\sim}{y}$

$$\hat{\underset{\sim}{b}} = (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{y},$$ for $\underset{\sim}{X}$ full column rank (e.g., regression)

$$\underset{\sim}{b}^{o} = (\underset{\sim}{X}'\underset{\sim}{X})^{-}\underset{\sim}{X}'\underset{\sim}{y},$$ otherwise (e.g., analysis of variance models)

GLS: $\underset{\sim}{X}'\underset{\sim}{R}^{-1}\underset{\sim}{X}\hat{\underset{\sim}{b}} = \underset{\sim}{X}'\underset{\sim}{R}^{-1}\underset{\sim}{y}$, for $\underset{\sim}{R}$ positive definite (p.d.)

$$\hat{\underset{\sim}{b}} = (\underset{\sim}{X}'\underset{\sim}{R}^{-1}\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{R}^{-1}\underset{\sim}{y},$$ for $\underset{\sim}{X}$ full column rank

$$= (\underset{\sim}{X}'\underset{\sim}{R}^{-1}\underset{\sim}{X})^{-}\underset{\sim}{X}'\underset{\sim}{R}^{-1}\underset{\sim}{y},$$ otherwise.

MIXED MODELS

<u>Model</u>: $\mathbf{y} = \mathbf{Xb} + \boldsymbol{\epsilon}$ with $\boldsymbol{\epsilon} = \mathbf{Zu} + \mathbf{e}$

$$= \mathbf{XB} + \mathbf{Zu} + \mathbf{e}$$

$\text{var}(\mathbf{e}) = \mathbf{R}$ (usually $\sigma^2\mathbf{I}$)

random effects: $\mathbf{u}' = [\mathbf{u}_1' \;\; \mathbf{u}_2' \cdots \mathbf{u}_c']$

effects of a random factor

$$\text{var}(\mathbf{u}) \overset{\text{def}}{=} \mathbf{D} = \text{diag}\left\{\sigma_1^2\mathbf{I}_{n_1} \quad \sigma_2^2\mathbf{I}_{n_2} \quad \cdots \quad \sigma_c^2\mathbf{I}_{n_c}\right\}$$

$$\text{var}(\mathbf{y}) \overset{\text{def}}{=} \mathbf{V} = \mathbf{ZDZ}' + \mathbf{R}$$

<u>Example</u>: y_{ijkt} = milk yield of the t'th cow, sired by sire k,

made in herd-year-season i at age j

$$= \mu + \underbrace{\underset{\text{h-y-s}}{h_i} + \underset{\text{age}}{a_j}}_{\text{fixed effects}} + \underbrace{\underset{\text{sire}}{s_k} + \underset{\text{cow}}{c_{kt}}}_{\text{random effects}} + e_{ijkt}$$

ESTIMATION IN MIXED MODELS

$$\underset{\sim}{y} = \underset{\sim}{Xb} + \underset{\sim}{Zu} + \underset{\sim}{e}$$

$$\underset{\sim}{V} = \text{var}(\underset{\sim}{y}) = \underset{\sim}{ZDZ'} + \underset{\sim}{R}$$

$\underset{\sim}{D}$ involves unknown variance components

Estimating fixed effects

GLS: $\underset{\sim}{X}'\underset{\sim}{V}^{-1}\underset{\sim}{X}\hat{\underset{\sim}{b}} = \underset{\sim}{X}'\underset{\sim}{V}^{-1}\underset{\sim}{y}$

Difficulty: $\underset{\sim}{V}$ involves unknown variance components.

They have to be estimated.

Estimation is usually from same data as used for $\hat{\underset{\sim}{b}}$.

Estimating variance components

(a) for their own sake

(b) for use in parameters defined in terms of variance components:

correlation, reliability and (in genetics) heritability, repeatability and genetic correlations

(c) for use in $\underset{\sim}{V}$ for estimating fixed effects

$$\underset{\sim}{X}'\tilde{\underset{\sim}{V}}^{-1}\underset{\sim}{X}\hat{\tilde{\underset{\sim}{b}}} = \underset{\sim}{X}'\tilde{\underset{\sim}{V}}^{-1}\underset{\sim}{y} .$$

PREDICTING REALIZED VALUES OF RANDOM EFFECTS

Example: For a dairy bull with N daughters having average first lactation yield $\bar{y}$: we use $\widehat{E(s|\bar{y})}$.

General case: $\underset{\sim}{y} = \underset{\sim}{X}\underset{\sim}{b} + \underset{\sim}{Z}\underset{\sim}{u} + \underset{\sim}{e}$ with $\underset{\sim}{V} = \text{var}(\underset{\sim}{y}) = \underset{\sim}{Z}\underset{\sim}{D}\underset{\sim}{Z}' + \underset{\sim}{R}$.

$$\widehat{E(\underset{\sim}{u}|\underset{\sim}{y})} = \underset{\sim}{D}\underset{\sim}{Z}'\underset{\sim}{V}^{-1}(\underset{\sim}{y} - \underset{\sim}{X}\hat{\underset{\sim}{b}}) \quad \text{with} \quad \underset{\sim}{X}'\underset{\sim}{V}^{-1}\underset{\sim}{X}\hat{\underset{\sim}{b}} = \underset{\sim}{X}'\underset{\sim}{V}^{-1}\underset{\sim}{y} .$$

Example: $y_{ij} = \mu + s_i + e_{ij}$, $j = 1, \cdots, n_i$, $i = 1, \cdots, c$: $\sigma_s^2 \equiv s$, $\sigma_e^2 \equiv e$.

Definitions:

$\underset{\sim}{I}_i = \underset{\sim}{I}_{n_i \times n_i}$, $\underset{\sim}{\mathbf{1}}'$ = vector of ones, $\underset{\sim}{J}_i$, $n_i \times n_i$, every element unity.

$$\underset{\sim}{D} = s\underset{\sim}{I}_{c\times c},\ \underset{\sim}{V} = \text{diag}\{e\underset{\sim}{I}_i + s\underset{\sim}{J}_i\},\ \underset{\sim}{V}^{-1} = \frac{1}{e}\,\text{diag}\{\underset{\sim}{I}_i - \lambda_i\underset{\sim}{J}_i\} \quad \text{with} \quad \lambda_i = \frac{n_i s}{e + n_i s} .$$

Estimation:

$$\hat{\mu} = \frac{\underset{\sim}{\mathbf{1}}'\underset{\sim}{V}^{-1}\underset{\sim}{y}}{\underset{\sim}{\mathbf{1}}'\underset{\sim}{V}^{-1}\underset{\sim}{\mathbf{1}}} = \frac{\Sigma(1 - \lambda_i)y_{i.}/e}{\Sigma(1 - \lambda_i)/e} = \Sigma\left(\frac{n_i}{e + n_i s}\right)\bar{y}_{i.} \Big/ \Sigma\left(\frac{n_i}{e + n_i s}\right) .$$

$$\widehat{E(\underset{\sim}{s}|\underset{\sim}{y})} = \underset{\sim}{D}\underset{\sim}{Z}'\underset{\sim}{V}^{-1}(\underset{\sim}{y} - \hat{\mu}) \Rightarrow \widehat{E(s_i|\underset{\sim}{y})} = s\,\frac{1}{e}(1 - \lambda_i)n_i(\bar{y}_{i.} - \hat{\mu}) = \lambda_i(\bar{y}_{i.} - \hat{\mu}).$$

Genetics: $$h = \frac{4\sigma_s^2}{\sigma_s^2 + \sigma_e^2} = \frac{4s}{s + e} = 4/(1 + \rho) \quad \text{for} \quad \rho = e/s .$$

$$\lambda_i = \frac{n_i}{\rho + n_i} = \frac{n_i}{4/h - 1 + n_i} = \frac{n_i h}{4 + (n_i - 1)h} .$$

Result: $$\widehat{E(s_i|\underset{\sim}{y})} = \frac{n_i h}{4 + (n_i - 1)h}\,(\bar{y}_{i.} - \hat{\mu}) .$$

METHODS OF ESTIMATING σ^2's

(Remainder of this paper.)

BALANCED DATA

Equate mean squares of ANOVA to expected values.

Example:

$$n_{ij}: \quad \begin{array}{|cccc|} 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 \end{array} \qquad y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}$$

Mean Squares	d.f.	Expected Mean Squares		
		Random Model	Mixed Model	Mixed Model with Σ-restrictions*
Rows	2	$8\sigma^2_\alpha + 2\sigma^2_\gamma + \sigma^2_e$		
Columns	3	$6\sigma^2_\beta + 2\sigma^2_\gamma + \sigma^2_e$	$6\sigma^2_\beta + 2\sigma^2_\gamma + \sigma^2_e$	$6\sigma^2_{\beta''} + \sigma^2_e$
Interaction	6	$2\sigma^2_\gamma + \sigma^2_e$	$2\sigma^2_\gamma + \sigma^2_e$	$2\sigma^2_\gamma + \sigma^2_e$
Error	12	σ^2_e	σ^2_e	σ^2_e

* Reference: "Linear Models", p. 403, $\sigma^2_{\beta''} = \sigma^2_\beta + \sigma^2_\gamma/3$.

Properties:

Easy to compute.

Unbiased estimators.

Minimum variance quadratic estimators and (under normality) minimum variance unbiased.

Under normality: sampling variances known (as functions of unknown σ^2's) and unbiased estimators available.

UNBALANCED DATA

At least 6 methods: 3 ANOVA and 3 ML ideas.

A useful computing device

$$\mathrm{tr}(\underset{\sim}{P}\underset{\sim}{Z}_i\underset{\sim}{Z}_i'\underset{\sim}{P}\underset{\sim}{Z}_j\underset{\sim}{Z}_j') = \mathrm{tr}\left[\underset{\sim}{Z}_i'\underset{\sim}{P}\underset{\sim}{Z}_j(\underset{\sim}{Z}_i'\underset{\sim}{P}\underset{\sim}{Z}_j)'\right]$$

$$= \sum_k \left\{\text{inner product of k'th row of } \underset{\sim}{Z}_i'\underset{\sim}{P}\underset{\sim}{Z}_j \text{ with itself}\right\}$$

$$= \text{sum of squares of elements of } \underset{\sim}{Z}_i'\underset{\sim}{P}\underset{\sim}{Z}_j \ .$$

Henderson's 3 methods: — based on ANOVA

E(MS) terms are linear functions of σ^2's.

Equate calculated MS terms to E(MS) terms.

Solve resulting linear equations.

Method 1: Uses mean squares similar to those of ANOVA for balanced data.

Easy to compute.

Not suitable for mixed models.

Method 2: Adjusts data for the fixed effects so that Method 1 can effectively be used for mixed models.

Moderately easy to compute.

Method 3: Uses fitting constants sums of squares.

Often hard (or impossible) to compute.

Suitable for mixed models.

Not uniquely defined.

Combinations of sums of squares available for Method 3 in the 2-way classification, random model.

	(i)	(ii)	(iii)
	$R(\alpha\|\mu)$	$R(\beta\|\mu)$	$R(\alpha\|\mu,\beta)$
	$R(\beta\|\mu,\alpha)$	$R(\alpha\|\mu,\beta)$	$R(\beta\|\mu,\alpha)$
	$R(\gamma\|\mu,\alpha,\beta)$	$R(\gamma\|\mu,\alpha,\beta)$	$R(\gamma\|\mu,\alpha,\beta)$
	SSE	SSE	SSE
SAS Output:	Type I, for A, B, AB	Type I, for B, A, AB	Type II
LM* Tables:	10.1, p. 447	10.2, p. 448	
LM p. 483, last 3 equations:	2nd and 3rd	1st and 3rd	1st and 2nd

* "Linear Models".

General difficulty with ANOVA-style methods:

No criteria, defined in terms of properties of estimators, upon which to choose one method over another; e.g., they all give unbiased estimators, distributions are unknown (even under normality), and negative estimates can occur.

MAXIMUM LIKELIHOOD

Notation: $\mathbf{V} = \mathbf{ZDZ}' + \sigma_e^2\mathbf{I} \qquad (\text{with } \mathbf{R} \equiv \sigma_e^2\mathbf{I})$

$$= \sum_{i=1}^{c} \mathbf{Z}_i\mathbf{Z}_i'\sigma_i^2 + \sigma_e^2\mathbf{I}$$

$$= \sum_{i=0}^{c} \mathbf{Z}_i\mathbf{Z}_i'\sigma_i^2 \qquad (\text{with } \sigma_0^2 \equiv \sigma_e^2,\ \mathbf{Z}_0 = \mathbf{I})$$

$$\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1} .$$

Distributional assumption: Normality.

Maximum Likelihood (ML)

c + 1 equations:

$$\left\{\mathrm{tr}(\tilde{\mathbf{V}}^{-1}\mathbf{Z}_i\mathbf{Z}_i'\tilde{\mathbf{V}}^{-1}\mathbf{Z}_j\mathbf{Z}_j')\right\}\left\{\tilde{\sigma}_i^2\right\} = \left\{\mathbf{y}'\mathbf{PZ}_i\mathbf{Z}_i'\mathbf{Py}\right\} , \text{ for } i,j = 0,\ 1,\ \cdots,\ c.$$

Solve by iteration: retain only positive values.

Large sample distributional properties known.

$$\tilde{\mathbf{b}} = (\mathbf{X}'\tilde{\mathbf{V}}^{-1}\mathbf{X})^{-}\mathbf{X}'\tilde{\mathbf{V}}^{-1}\mathbf{y} \qquad \tilde{\mathbf{u}} = \mathbf{DZ}'\tilde{\mathbf{V}}^{-1}(\mathbf{y} - \mathbf{X}\tilde{\mathbf{b}}) .$$

Restricted Maximum Likelihood (REML)

c + 1 equations:

$$\left\{\mathrm{tr}(\mathbf{PZ}_i\mathbf{Z}_i'\mathbf{PZ}_j\mathbf{Z}_j')\right\}\left\{\sigma_i^2\right\} = \left\{\mathbf{y}'\mathbf{PZ}_i\mathbf{Z}_i'\mathbf{Py}\right\} , \text{ for } i,j = 0,\ 1,\ \cdots,\ c.$$

Similar properties to ML except REML is the same as ANOVA, for balanced data.

MINQUE: Minimum Norm Quadratic Unbiased Estimation

No distribution assumption needed.

No iteration.

Requires assigning à priori values ($\underset{\sim}{w}$, say) to σ^2's.

$\underset{\sim}{D}_w$, $\underset{\sim}{V}_w$, $\underset{\sim}{P}_w$ are $\underset{\sim}{D}$, $\underset{\sim}{V}$ and $\underset{\sim}{P}$ with w's in place of σ^2's.

Solve c + 1 equations:

$$\left\{\mathrm{tr}(\underset{\sim}{P}_w\underset{\sim}{Z}_i\underset{\sim}{Z}_i'\underset{\sim}{P}_w\underset{\sim}{Z}_j\underset{\sim}{Z}_j')\right\}\left\{\sigma_i^2\right\} = \left\{\underset{\sim}{y}'\underset{\sim}{P}_w\underset{\sim}{Z}_i\underset{\sim}{Z}_i'\underset{\sim}{P}_w\underset{\sim}{y}\right\}, \text{ for } i,j = 0, 1, \cdots, c.$$

Note: $\underset{\sim}{w}$ is numeric and so is $\underset{\sim}{P}_w$.

Estimator depends on $\underset{\sim}{w}$.

There are many MINQUE's.

(a) REML and MINQUE

REML involves iteration but MINQUE does not. Yet equation systems are essentially the same, except that MINQUE involves an à priori value $\underset{\sim}{w}$ in $\underset{\sim}{P}$ in place of $\underset{\sim}{\sigma}^2$. If that $\underset{\sim}{w}$ were used as the starting value for the iterations of REML, the resulting first iterate would be the MINQUE based on $\underset{\sim}{w}$;

i.e., 1st iterate of REML = a MINQUE .

(b) Iterative MINQUE

Having got the MINQUE, $\underset{\sim}{\sigma}^2_{(w)}$, based on $\underset{\sim}{w}$, now use it in place of $\underset{\sim}{w}$; and keep going, i.e., iterate MINQUE. This gives REML;

i.e., Iterative MINQUE = REML .

(c) MINQUE(0) (or MIVQUE0)

Special case of MINQUE: i.e., a particular $\underset{\sim}{w}$.

$\underset{\sim}{w}' = [\sigma_e^2 = 1$, all other σ^2's = zero]; $\underset{\sim}{P}_w = \underset{\sim}{I} - \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^-\underset{\sim}{X}$.

SAS VARCOMP: ingenious computing, no matrix inversion, and feasible even for very large data sets.

But remember, MINQUE0 is based on à priori values of zero for all variance components save σ_e^2. In many cases this may not be at all realistic.

A SHORT REFERENCE LIST

The following references serve as an entrée to the literature on variance components estimation.

Analysis of variance methods

Henderson, C. R. [1953]. Estimation of variance and covariance components. Biometrics 9:226-252.

Henderson, C. R., Searle, S. R. and Schaeffer, L. R. [1974]. The invariance and calculation of Method 2 for estimating variance components. Biometrics 30: 583-588.

Searle, S. R. [1968]. Another look at Henderson's methods of estimating variance components. Biometrics 24:749-787.

Searle, S. R. [1971]. Linear Models. Wiley, New York. (Chapters 9-11)

Maximum Likelihood

Corbeil, R. R. and Searle, S. R. [1976]. A comparison of variance components estimators. Biometrics 32:779-791.

Hartley, H. O. and Rao, J. N. K. [1967]. Maximum likelihood estimation for the mixed analysis of variance model. Biometrika 54:93-108.

Harville, D. A. [1977]. Maximum likelihood approaches to variance component estimation and to related problems. J. Am. Stat. Assoc. 72:320-340.

Jennrich, R. I. and Sampson, P. F. [1976]. Newton-Raphson and related algorithms for maximum likelihood estimation of variance components. Technometrics 18: 11-18.

Miller, J. J. [1973]. Asymptotic properties and computation of maximum likelihood estimates in the mixed model of the analysis of variance. Technical Report 12, Dept. of Statistics, Stanford University, Stanford, California.

Searle, S. R. [1979]. Notes on variance component estimation: a detailed account of maximum likelihood and kindred methodology. Report No. BU-673-M, Biometrics Unit, Cornell University.

MINQUE

Goodnight, J. H. [1978]. Computing MIVQUE0 estimates of variance components. Technical Report R-105, SAS Institute, Raleigh, North Carolina.

Rao, C. R. [1972]. Estimation of variance and covariance components in linear models. J. Am. Stat. Assoc. 67:112-115.

VanVleck, L. Dale and Searle, Shayle R. (Eds.), [1979]. Proceedings of Variance Components and Animal Breeding: A Conference in Honor of C. R. Henderson. Animal Science Department, Cornell University. (Especially papers by Anderson, Brown, Pukelsheim and Quaas.)

DATA SET 1

This is a composite page.

VARIANCE COMPONENT ESTIMATION PROCEDURE

CLASS LEVEL INFORMATION

BALANCED DATA

CLASS	LEVELS	VALUES
A	3	1 2 3
B	4	1 2 3 4

NUMBER OF OBSERVATIONS IN DATA SET = 24

RANDOM MODEL

SOURCE	DF	TYPE I SS	TYPE I MS	EXPECTED MEAN SQUARE
A	2	448.00000000	224.00000000	VAR(ERROR) + 2 VAR(A*B) + 8 VAR(A)
B	3	36.00000000	12.00000000	VAR(ERROR) + 2 VAR(A*B) + 6 VAR(B)
A*B	6	136.00000000	22.66666667	VAR(ERROR) + 2 VAR(A*B)
ERROR	12	250.00000000	20.83333333	VAR(ERROR)
CORRECTED TOTAL	23	870.00000000		

These are actually Henderson Method III estimates (see page 5), but are the same as ANOVA estimates because the data are balanced.

VARIANCE COMPONENT	ESTIMATE	
VAR(A)	25.16666667	$= (224 - 22\frac{2}{3})/8 = 25\frac{1}{6}$
VAR(B)	-1.77777778	$= (12 - 22\frac{2}{3})/6 = -1\frac{7}{9}$ ← This negative value plays a role in maximum likelihood estimation (see page 9).
VAR(A*B)	0.91666667	$= (22\frac{2}{3} - 20\frac{5}{6})/2 = 11/12$
VAR(ERROR)	20.83333333	

MIXED MODEL

(rows as fixed effects)

SOURCE	DF	TYPE I SS	TYPE I MS	EXPECTED MEAN SQUARE
A	2	448.00000000	224.00000000	VAR(ERROR) + 2 VAR(A*B) + $\boxed{\text{G(A)}}$
B	3	36.00000000	12.00000000	VAR(ERROR) + 2 VAR(A*B) + 6 VAR(B)
A*B	6	136.00000000	22.66666667	VAR(ERROR) + 2 VAR(A*B)
ERROR	12	250.00000000	20.83333333	VAR(ERROR)
CORRECTED TOTAL	23	870.00000000		

← A quadratic function in the effects for Factor A. See LM 401.

VARIANCE COMPONENT	ESTIMATE
AR(B)	-1.77777778
·) (A*B)	0.91666667
ERROR)	20.83333333

Same as for random model, because data are balanced.

11

I

DATA SET 1

MIVQUE(0) VARIANCE COMPONENT ESTIMATION PROCEDURE

SSQ MATRIX

RANDOM MODEL

SOURCE	A	B	A*B	ERROR	Y	
A	128.00000000	0.00000000	32.00000000	16.00000000	3584.00000000	$= b^2n^2\Sigma(\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot\cdot\cdot})^2$
B	0.00000000	108.00000000	36.00000000	18.00000000	216.00000000	$= a^2n^2\Sigma(\bar{y}_{\cdot j\cdot} - \bar{y}_{\cdot\cdot\cdot})^2$
A*B	32.00000000	36.00000000	44.00000000	22.00000000	1240.00000000	$= n^2\Sigma\Sigma(\bar{y}_{ij\cdot} - \bar{y}_{\cdot\cdot\cdot})^2$
ERROR	16.00000000	18.00000000	22.00000000	23.00000000	870.00000000	$= SST_m$

VARIANCE COMPONENT	ESTIMATE Y	
VAR(A)	25.16666667	Same as ANOVA estimates on page I, because the data are balanced.
VAR(B)	-1.77777778	
VAR(A*B)	0.91666667	
VAR(ERROR)	20.83333333	

MIVQUE(0) estimation assumes normality. It is the same as MINQUE (which requires no distributional assumptions) with the pre-assigned value of var($\underset{\sim}{y}$) being taken as $\underset{\sim}{I}$. The basic equations for this are based on writing the linear model as

$$\underset{\sim}{y} = \underset{\sim}{X}\underset{\sim}{\alpha} + \sum_{i=1}^{c+1} \underset{\sim}{Z}_i \underset{\sim}{b}_i$$

where $\underset{\sim}{\alpha}$ represents fixed effects and the $\underset{\sim}{b}_i$ for $i = 1, \cdots, c+1$ are vectors of the effects for each random factor, with $\underset{\sim}{b}_{c+1} = \underset{\sim}{e}$, the random errors, and $\underset{\sim}{Z}_{c+1} = \underset{\sim}{I}$. Then, for $\underset{\sim}{\sigma}^2$ being the vector of variance components to be estimated, including the error variance, $\underset{\sim}{\sigma}^2_{c+1}$, the estimation equations are

$$\{tr(\underset{\sim}{M}\underset{\sim}{Z}_i\underset{\sim}{Z}_i'\underset{\sim}{M}\underset{\sim}{Z}_j\underset{\sim}{Z}_j')\}\hat{\underset{\sim}{\sigma}}^2 = \{\underset{\sim}{y}'\underset{\sim}{M}\underset{\sim}{Z}_i\underset{\sim}{Z}_i'\underset{\sim}{M}\underset{\sim}{y}\} \quad \text{for } i,j = 1, \cdots, c, c+1,$$

where

$$\underset{\sim}{M} = \underset{\sim}{I} - \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-}\underset{\sim}{X}' .$$

These equations can be written in equivalent form as

$$\{\text{s/s elements of } \underset{\sim}{Z}_i'\underset{\sim}{M}\underset{\sim}{Z}_j\}\hat{\underset{\sim}{\sigma}}^2 = \{\text{s/s of elements of } \underset{\sim}{Z}_i'\underset{\sim}{M}\underset{\sim}{y}\} \quad \text{for } i = 1, \cdots, c+1.$$

The above output is of this form:

$$\begin{bmatrix} 128 & 0 & 32 & 16 \\ 0 & 108 & 36 & 18 \\ 32 & 36 & 44 & 22 \\ 16 & 18 & 22 & 23 \end{bmatrix} \begin{bmatrix} 25\frac{1}{6} \\ -1\frac{7}{9} \\ \frac{11}{12} \\ 20\frac{10}{12} \end{bmatrix} = \begin{bmatrix} 3584 \\ 216 \\ 1240 \\ 870 \end{bmatrix} . \qquad (P2)$$

12

Derivation of the terms in these equations is illustrated on page 5.

DATA SET 1

MAXIMUM LIKELIHOOD VARIANCE COMPONENT ESTIMATION PROCEDURE

EPENDENT VARIABLE: Y

RANDOM MODEL

ITERATION	OBJECTIVE	VAR(A)	VAR(B)	VAR(A*B)	VAR(ERROR)
0	78.21992029	22.63766340	0.00000019	0.82455065	18.73978758
1	78.06631159	14.43042667	0.00000000	0.42733352	20.03304904
2	78.03224210	16.04623292	0.00000000	0.07684525	20.04761726
3	78.02882715	16.20253102	0.00000000	0.00302434	20.08551836
4	78.02868361	16.16334896	0.00000000	0.00000053	20.09394261
5	78.02868324	16.15583146	0.00000000	0.00000000	20.09507672
6	78.02868323	16.15489553	0.00000000	0.00000000	20.09521793

CONVERGENCE CRITERION MET

The convergence criterion is that two successive values of the $\log|\underset{\sim}{V}|$ (the objective function) differ by less than 0.00000001.

The equations for maximum likelihood estimation of variance components from unbalanced data have to be solved iteratively. For some cases of balanced data (e.g., see Corbeil and Searle, Biometrics, 32, 779, 1976), the equations can be solved analytically. This is not so for the 2-way crossed classification random model (e.g., J. J. Miller, Technical Report 12, Stanford University, 1973). When iteration is required, the SAS Manual indicates that it begins with using the MIVQUE(0) estimates as the initial values.

In all cases of maximum likelihood estimation of variance components, solutions of the equations have to be confined to positive values (the parameter space) in order to be estimates. This is usually handled by truncating negative values to zero – and by keeping them zero in successive iterations.

Truncation of this sort has occurred here, for $\hat{\sigma}^2_\beta$ (with starting value -16/9 from page 1) and for $\hat{\sigma}^2_{\alpha\beta}$. This effectively reduces the case to the 1-way classification model

$$y_{it} = \mu + \alpha_i + e_{it} \quad \text{for } i = 1, \cdots, 3 = a' \text{ and } t = 1, \cdots, 8 = n', \text{ with } a' = a \text{ and } n' = bn.$$

Accordingly (Corbeil and Searle, loc. cit., Table 1) the estimators are, for this 1-way model, with SSE' = SSE + SSAB + SSB and SSA' = SSA (see top of page 10

$$\hat{\sigma}^2_e = \frac{\text{SSE for this 1-way model}}{a'(n'-1)} = \frac{36 + 136 + 250}{3(7)} = \frac{422}{21} = 20\tfrac{2}{21} = \underline{20.095}2381$$

and

$$\hat{\sigma}^2_\alpha = \frac{1}{n'}\left(\frac{SSA'}{a'} - \hat{\sigma}^2_e\right) = \frac{1}{8}\left(\frac{448}{3} - 20\tfrac{2}{21}\right) = \frac{1}{8}\left(129\tfrac{5}{21}\right) = 16\tfrac{13}{84} \qquad = \underline{16.154}7619 .$$

To 4 decimal places, these are the solutions in the above output.

The objective function is $\log|\underset{\sim}{V}|$ – see page 10. With $\sigma^2_\beta = 0$ and $\sigma^2_\gamma = 0$, we have $\underset{\sim}{V} = \underset{\sim}{I}_3 \otimes (\sigma^2_e \underset{\sim}{I}_8 + \sigma^2_\alpha \underset{\sim}{J}_8)$ and so $|\underset{\sim}{V}| = [(\sigma^2_e)^7(\sigma^2_e + 8\sigma^2_\alpha)]^3$. Hence, for the ML estimates, the objective function is

$$\log|\tilde{V}| = 21 \log\tilde{\sigma}^2_e + 3\log(\tilde{\sigma}^2_e + 8\tilde{\sigma}^2_\alpha) = 21 \log_e 20.0952 + 3 \log_e[20.0952 + 8(16.1549)] = 78.028, \text{ as shown.}$$

13

DATA SET 2NB (NESTED, BALANCED)

MAXIMUM LIKELIHOOD VARIANCE COMPONENT ESTIMATION PROCEDURE

DEPENDENT VARIABLE: Y

ITERATION	OBJECTIVE	VAR(A)	VAR(B(A))	VAR(ERROR)
0	33.73133446	144.83333333	5.42361111	5.72916667
1	36.23441597	20.23334221	10.61902943	10.18124689
2	43.76701889	101.83288952	932.40300711	4.27422416
3	49.80718444	378.08444592	4491.55495835	4.18945934
4	55.44000184	1502.43695048	18572.88237597	4.17218644

SERIES DIVERGED, PROCEDURE TERMINATED.

This is an error.

The data are balanced, and the ML equations have exact solutions – see page 14. As of May 6, 1980, this error is being investigated by SAS personnel. Two features of this investigation are as follows.

(i) When initial values $\sigma^2_\alpha = \sigma^2_{\beta:\alpha} = \sigma^2_e = s$, say are used (i.e., each component given the same initial value s), then for s in the range 1 through 3.3, the resulting estimates are

$$\tilde{\sigma}^2_\alpha = 0, \qquad \tilde{\sigma}^2_{\beta:\alpha} = 82.915927 \quad \text{and} \quad \tilde{\sigma}^2_e = 6.25 .$$

The error seems to be related to the definition of the nesting. Having put $\tilde{\sigma}^2_\alpha = 0$, the other estimates are correct, because $\sigma^2_\alpha = 0$ reduces the model to $y_{kt} = \mu + \beta_k + e_{kt}$ for $k = 1, \cdots, 4$ and $t = 1, 2, 3$, with the following analysis of variance:

Term	d.f.	S.S.	M.S.	E(MS)	Estimates: ANOVA	Estimates: ML
Mean	1	3072				
Groups	3	1020	340	$\sigma^2_e + 3\sigma^2_\beta$	$\hat{\sigma}^2_\beta = 333\frac{3}{4}/3 = 111\frac{1}{4}$	$\tilde{\sigma}^2_\beta = \frac{1}{3}\left[\frac{1020}{3} - 6\frac{1}{4}\right] = 82\frac{11}{12} = 82.91\dot{6}$
Error	8	50	$6\frac{1}{4}$	σ^2_e	$\hat{\sigma}^2_e = 6\frac{1}{4}$	$\tilde{\sigma}^2_e = 6\frac{1}{4}$
Total	12	4142				

14

DATA SET 2NU (NESTED, UNBALANCED)

This is a composite page.

VARIANCE COMPONENT ESTIMATION PROCEDURE

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
A	2	1 2
B	3	1 2 3

DATA SET 2NU
(nested, unbalanced)

NUMBER OF OBSERVATIONS IN DATA SET = 12

DEPENDENT VARIABLE: Y

SOURCE	DF	TYPE I SS	TYPE I MS	EXPECTED MEAN SQUARE (See pages 15-17.)
A	1	864.00000000	864.00000000	VAR(ERROR) + 2.66666667 VAR(B(A)) + 5.33333333 VAR(A)
B(A)	3	60.00000000	20.00000000	VAR(ERROR) + 2.16666667 VAR(B(A))
ERROR	7	152.00000000	21.71428571	VAR(ERROR)
CORRECTED TOTAL	11	1076.00000000		

VARIANCE COMPONENT	ESTIMATE
VAR(A)	158.32417582
VAR(B(A))	-0.79120879
VAR(ERROR)	21.71428571

Henderson Method III estimates, equivalent to ANOVA estimates (see page 15), because the data are nested.

MIVGUE(0) VARIANCE COMPONENT ESTIMATION PROCEDURE

SSQ MATRIX

SOURCE	A	B(A)	ERROR	Y
A	28.44444444	14.22222222	5.33333333	4608.00000000
B(A)	14.22222222	24.02777778	9.16666667	2786.00000000
ERROR	5.33333333	9.16666667	11.00000000	1076.00000000

VARIANCE COMPONENT	ESTIMATE Y
VAR(A)	147.77019041
VAR(B(A))	27.12015758
VAR(ERROR)	3.57189757

Derivation is shown on pages 18-19.

DATA SET 2

This is a composite page.

VARIANCE COMPONENT ESTIMATION PROCEDURE

DEPENDENT VARIABLE: Y

RANDOM MODEL

SOURCE	DF	TYPE I SS	TYPE I MS	EXPECTED MEAN SQUARE (See pages 23-24 and 26-29 for derivation.)
A	3	$R(\alpha\mid\mu)$ = 162.00000000	54.00000000	VAR(ERROR) + 0.25925926 VAR(B) + 2.14814815 VAR(A)
B	2	$R(\beta\mid\mu,\alpha)$ = 258.00000000	129.00000000	VAR(ERROR) + 2.5 VAR(B)
ERROR	3	SSE = 12.00000000	4.00000000	VAR(ERROR)
CORRECTED TOTAL	8	432.00000000		

This is Henderson Method III estimation.

VARIANCE COMPONENT	ESTIMATE
VAR(A)	[54 - 4 - .259259(50)]/2.148148 = 17.24137931
VAR(B)	(129 - 4)/2.5 = 50.00000000
VAR(ERROR)	4.00000000

Based on $R(\mu,\alpha,\beta)$ and $R(\mu,\alpha)$ but not $R(\mu,\beta)$.

This is sub-method (i) as described on page 25.

MIXED MODEL

Factor A (rows) is fixed.

SOURCE	DF	TYPE I SS	TYPE I MS	EXPECTED MEAN SQUARE
A	3	162.00000000	54.00000000	VAR(ERROR) + 0.25925926 VAR(B) + $\boxed{\text{G(A)}}$
B	2	258.00000000	129.00000000	VAR(ERROR) + 2.5 VAR(B)
ERROR	3	12.00000000	4.00000000	VAR(ERROR)
CORRECTED TOTAL	8	432.00000000		

VARIANCE COMPONENT	ESTIMATE
VAR(B)	50.00000000
VAR(ERROR)	4.00000000

With rows as fixed effects, only $R(\beta\mid\mu,\alpha)$ and SSE are used. This is equivalent to sub-method (i) on page 25, ignoring $R(\alpha\mid\mu)$.

DATA SET 2

This is a composite page.

MAXIMUM LIKELIHOOD VARIANCE COMPONENT ESTIMATION PROCEDURE

DEPENDENT VARIABLE: Y

RANDOM MODEL

See comments on page 20.

ITERATION	OBJECTIVE	VAR(A)	VAR(B)	VAR(ERROR)
0	34.10984587	12.76471432	14.11478621	23.84129115
1	33.92056173	15.43032249	15.71638555	20.74077187
2	33.71636021	17.71801586	17.65145797	17.94960474
3	33.49921369	19.87117758	19.56612501	15.48099805
4	33.27620560	21.92084362	21.38903554	13.34137889
5	33.05622572	23.86697025	23.10566489	11.52341287
6	32.84647013	25.70942658	24.72003845	10.00630058
7	32.66102672	27.45099858	26.24162646	8.75993548
8	32.49977193	29.09377475	27.67772552	7.74981631
9	32.36773936	30.63510458	29.02897202	6.94100698
10	32.26502246	32.06571235	30.28800277	6.30065174
11	32.18918674	33.37087498	31.44119682	5.79922997
12	32.13606646	34.53427231	32.47278632	5.41095402
13	32.10072643	35.54302383	33.36987295	5.11369249
14	32.07833714	36.39191365	34.12652896	4.88869491
15	32.06477310	37.08523703	34.74560040	4.72027345
16	32.05667292	37.63589791	35.23792783	4.59549844
17	32.05242261	38.06259165	35.61978626	4.50390535
18	32.04998321	38.38643254	35.90980175	4.43719456
19	32.04867473	38.62812036	36.12635516	4.38891698
20	32.04798452	38.80614017	36.28592020	4.35415604
21	32.04762502	38.93595441	36.40230763	4.32922499
22	32.04743953	39.02995763	36.48655908	4.31179669
23	32.04734449	39.09753024	36.54720703	4.29867543
24	32.04729603	39.14600476	36.59068602	4.28961278
25	32.04727141	39.18065143	36.62176431	4.28316404
26	32.04725894	39.20536257	36.64393144	4.27857913
27	32.04725263	39.22296069	36.65971839	4.27532133
28	32.04724945	39.23547976	36.67094924	4.27300751
29	32.04724784	39.24437865	36.67893265	4.27136463
30	32.04724703	39.25070085	36.68460449	4.27019841
31	32.04724662	39.25519067	36.68863248	4.26937068
32	32.04724642	39.25837831	36.69149224	4.26878325
33	32.04724631	39.26064097	36.69352219	4.26836641
34	32.04724626	39.26224685	36.69496291	4.26807061
35	32.04724623	39.26338648	36.69598533	4.26786073
36	32.04724622	39.26419517	36.69671084	4.26771182
37	32.04724621	39.26476911	36.69722565	4.26760616

CONVERGENCE CRITERION MET

MIXED MODEL

(rows are fixed effects)

ITERATION	OBJECTIVE	VAR(B)	VAR(ERROR)
	18.14311644	25.79795396	2.34526854
1	18.01410477	29.07768125	2.16942395
2	17.97460193	31.44402745	2.07553684
3	17.96577168	32.76923744	2.03158057
4	17.96433342	33.35877063	2.01363629
5	17.96410288	33.58498339	2.00699071
6	17.964078[illegible]	33.66579954	2.00464657
7	17.96407577	33.69386257	2.00383737
8	17.96407468	33.70350705	2.00355912
9	17.96407464	33.70680972	2.00346401
10	17.96407464	33.70793930	2.00343148

CONVERGENCE CRITERION MET

17

XIII

DATA SET 3

This is a composite page.

VARIANCE COMPONENT ESTIMATION PROCEDURE

RANDOM MODEL, WITH INTERACTION

DEPENDENT VARIABLE: Y

This is Henderson Method III estimation.

SOURCE		DF	TYPE I SS	TYPE I MS	EXPECTED MEAN SQUARE
A	$R(\alpha\|\mu)$	1	8.00000000	8.00000000	VAR(ERROR) + 1.5 VAR(A*B) + 0.25 VAR(B) + 4 VAR(A)
B	$R(\beta\|\mu,\alpha)$	2	11.63636364	5.81818182	VAR(ERROR) + 1.28787879 VAR(A*B) + 2.5 VAR(B)
A*B	$R(\gamma\|\mu,\alpha,\beta)$	2	36.36363636	18.18181818	VAR(ERROR) + 1.21212121 VAR(A*B)
ERROR	SSE	2	10.00000000	5.00000000	VAR(ERROR)
CORRECTED TOTAL	SST_m	7	66.00000000		

Derivation is shown on pages 35-38.

VARIANCE COMPONENT	ESTIMATE	This is Henderson Method III estimation	
VAR(A)	-2.99843750	$= \frac{1}{4}[8 - 5 - 1\frac{1}{2}(10\frac{7}{8}) - \frac{1}{4}(-5\frac{11}{40})]$	$= -2\frac{159}{160}$
VAR(B)	-5.27500000	$= [5\frac{9}{11} - 5 - 1\frac{19}{66}(10\frac{7}{8})]/2.5$	$= -5\frac{11}{40}$
VAR(A*B)	10.87500000	$= (18\frac{2}{11} - 5)/(13\frac{1}{3}/11)$	$= 10\frac{7}{8}$
VAR(ERROR)	5.00000000		

Note: This is Henderson Method III estimation using SAS GLM Type I sums of squares for fitting the factors in the sequence A, B and AB. The same estimation method could be used also on SAS GLM Type I sums of squares for the sequence B, A, AB, or on SAS GLM Type II sums of squares. The estimates would not be the same in all three cases. Correspondence with LM is as follows:

SAS	LM
Type I for A, B, AB	Table 10.1, LM 447
Type I for B, A, AB	Table 10.2, LM 448
Type II	3rd and 2nd to last equations on LM 483.

DEPENDENT VARIABLE: Y

MIXED MODEL

(rows are fixed effects)

SOURCE	DF	TYPE I SS	TYPE I MS	EXPECTED MEAN SQUARE
A	1	8.00000000	8.00000000	VAR(ERROR) + 1.5 VAR(A*B) + 0.25 VAR(B) + Q(A)
B	2	11.63636364	5.81818182	VAR(ERROR) + 1.28787879 VAR(A*B) + 2.5 VAR(B)
A*B	2	36.36363636	18.18181818	VAR(ERROR) + 1.21212121 VAR(A*B)
ERROR	2	10.00000000	5.00000000	VAR(ERROR)
CORRECTED TOTAL	7	66.00000000		

VARIANCE COMPONENT	ESTIMATE
VAR(B)	-5.27500000
VAR(A*B)	10.87500000
VAR(ERROR)	5.00000000

In the random model, σ_e^2, σ_γ^2 and σ_β^2 are estimated by using SSE, $R(\gamma|\mu,\alpha,\beta)$ and $R(\beta|\mu,\alpha)$, respectively — as above. These are unaltered in the mixed model, with factor A taken as fixed. These estimates are therefore identical to those of the random model.

DATA SET 3

This is a composite page.

MIVQUE(0) VARIANCE COMPONENT ESTIMATION PROCEDURE

SSQ MATRIX

RANDOM MODEL

SOURCE	A	B	A*B	ERROR	Y
A	16.00000000	1.00000000	6.00000000	4.00000000	32.00000000
B	1.00000000	14.06250000	7.62500000	5.25000000	18.00000000
A*B	6.00000000	7.62500000	9.25000000	6.50000000	60.00000000
ERROR	4.00000000	5.25000000	6.50000000	7.00000000	66.00000000

VARIANCE COMPONENT	ESTIMATE Y
VAR(A)	-2.01191827
VAR(B)	-4.93530079
VAR(A*B)	5.25312145
VAR(ERROR)	9.40181612

These are different from Henderson Method III estimates (page XIV) because the data are unbalanced.

SSQ MATRIX

MIXED MODEL

SOURCE	B	A*B	ERROR	Y
B	12.62500000	6.50000000	5.00000000	32.00000000
A*B	6.50000000	6.50000000	5.00000000	64.00000000
ERROR	5.00000000	5.00000000	6.00000000	58.00000000

VARIANCE COMPONENT	ESTIMATE Y
VAR(B)	-5.22448980
VAR(A*B)	11.93877551
VAR(ERROR)	4.07142857

These are different from MIVQUE(0) for the random model (as above) and from Henderson Method III estimates (page XIV), because the data are unbalanced.

19

DATA SET 3

This is a composite page.

MAXIMUM LIKELIHOOD VARIANCE COMPONENT ESTIMATION PROCEDURE

DEPENDENT VARIABLE: Y

RANDOM MODEL

ITERATION	OBJECTIVE	VAR(A)	VAR(B)	VAR(A*B)	VAR(ERROR)
0	17.13793086	0.00000006	0.00000006	3.16030420	5.65617970
1	17.08626906	0.00000000	0.00000000	2.54742193	6.11005709
2	17.03954767	0.00000000	0.00000000	1.95732378	6.57165287
3	16.99410785	0.00000000	0.00000000	1.37174979	7.05139639
4	16.94864981	0.00000000	0.00000000	0.79645601	7.54195395
5	16.90746876	0.00000000	0.00000000	0.29792862	7.98136399
6	16.88458551	0.00000000	0.00000000	0.03274041	8.22026250
7	16.88171203	0.00000000	0.00000000	0.00007289	8.24993374
8	16.88170560	0.00000000	0.00000000	0.00000000	8.25000000
9	16.88170560	0.00000000	0.00000000	0.00000000	8.25000000

CONVERGENCE CRITERION MET

Starting values of $\tilde{\sigma}^2_\alpha$ and $\tilde{\sigma}^2_\beta$ are negative (page XV) and are truncated to zero.

When $\tilde{\sigma}^2_\alpha$, $\tilde{\sigma}^2_\beta$ and $\tilde{\sigma}^2_\gamma$ have all been set to zero, the ML estimate of σ^2_e is

$$\tilde{\sigma}^2_e = SST_m/8 = (458 - 392)/8 = 66/8 = 8.25 .$$

MIXED MODEL

(rows are fixed effects)

ITERATION	OBJECTIVE	VAR(B)	VAR(A*B)	VAR(ERROR)
0	16.65859268	0.00000002	7.32951128	2.49955128
1	15.80181693	0.00000001	4.18520679	3.65379027
2	15.72983676	0.00000000	3.63642962	3.98398155
3	15.70014467	0.00000000	3.30099685	4.20868747
4	15.68621949	0.00000000	3.07614816	4.36908275
5	15.67923969	0.00000000	2.91859634	4.48610758
6	15.67559836	0.00000000	2.80539591	4.57252343
7	15.67364877	0.00000000	2.72279251	4.63680159
8	15.67258639	0.00000000	2.66190426	4.68483547
9	15.67200032	0.00000000	2.61671477	4.72[illegible]64094
10	15.67167415	0.00000000	2.58301670	4.74788689
11	15.67149148	0.00000000	2.557803[illegible]1	4.76823262
12	15.67138869	0.00000000	2.53889173	4.78355396
13	15.67133066	0.00000000	2.52468243	4.79510035
14	15.67129781	[illegible]	2.51399218	4.80380662
15	15.67127921	0.00000005	2.50594173	4.81037399
16	15.67126860	0.00000000	2.49987492	4.81532942
17	15.67126258	0.00000000	2.49530052	4.819[illegible]6937
18	15.67125916	0.00000000	2.49185005	4.82189243
19	15.67125721	0.00000000	2.48924657	4.824[illegible]2365
20	15.67125610	0.00000000	2.48728174	4.82563273
21	15.67125547	0.00000000	2.48579863	4.82684768
22	15.67125511	0.00000000	2.48467901	4.82776507
23	15.67125490	[illegible]	2.4838337[illegible]	4.82845782
24	15.67125482	0.0000000[illegible]	2.48319545	4.82898094
25	15.67125472	0.00000000	2.48271353	4.82937598
26	15.67125468	0.00000000	2.48234962	4.82967431
27	15.67125466	0.00000000	2.48207481	4.8298996[illegible]
28	15.67125465	0.00000000	2.48186728	4.83006974
29	15.67125464	0.00000000	2.48171057	4.83019823

CONVERGENCE CRITERION MET

Even though $\sigma^2_\beta = 0$ reduces this to a 2-way nested model

$$\mu + \alpha_i + \gamma_{j:i} + e_{ijk},$$

the data being unbalanced means that there is no simple structure to the ML equations.

20

DATA SET 4

This is a composite page.

MIVQUE(0) VARIANCE COMPONENT ESTIMATION PROCEDURE

SSQ MATRIX

RANDOM MODEL

SOURCE	A	B	A*B	ERROR	Y
A	23.04000000	0.64000000	10.24000000	4.80000000	288.00000000
B	0.64000000	21.64000000	11.64000000	5.80000000	6.50000000
A*B	10.24000000	11.64000000	16.44000000	7.80000000	141.50000000
ERROR	4.80000000	5.80000000	7.80000000	9.00000000	88.50000000

VARIANCE COMPONENT	ESTIMATE Y
VAR(A)	11.44060743
VAR(B)	-1.43562637
VAR(A*B)	0.48919241
VAR(ERROR)	4.23289072

Different from Henderson Method III estimates (page XVII) because the data are unbalanced.

SSQ MATRIX

MIXED MODEL

(rows are fixed effects)

SOURCE	B	A*B	ERROR	Y
B	21.11111111	11.44444444	5.66666667	0.50000000
A*B	11.44444444	11.44444444	5.66666667	5.50000000
ERROR	5.66666667	5.66666667	8.00000000	28.50000000

VARIANCE COMPONENT	ESTIMATE Y
VAR(B)	-0.51724138
VAR(A*B)	-1.45939413
VAR(ERROR)	4.96261682

Different from MIVQUE(0) for random model (as above) and from Henderson Method III estimates (page XVII), because the data are unbalanced.

MAXIMUM LIKELIHOOD VARIANCE COMPONENT ESTIMATION PROCEDURE

DEPENDENT VARIABLE: Y

RANDOM MODEL

ITERATION	OBJECTIVE	VAR(A)	VAR(B)	VAR(A*B)	VAR(ERROR)
0	17.62082437	8.69299492	0.00000003	0.37170641	3.21630628
1	17.10057907	4.16621993	0.00000001	0.02054055	3.80341643
2	17.01621660	4.97235006	0.00000000	0.00012936	3.64520224
3	17.00710868	5.34866015	0.00000000	0.00000005	3.58520639
4	17.00639076	5.46872045	0.00000000	0.00000000	3.56795019
5	17.00635414	5.49682449	0.00000000	0.00000000	3.56402381
6	17.00635257	5.50268820	0.00000000	0.00000000	3.56320985
7	17.00635251	5.50387720	0.00000000	0.00000000	3.56304502
8	17.00635250	5.50411685	0.00000000	0.00000000	3.56301180

CONVERGENCE CRITERION MET

MIXED MODEL

(rows are fixed effects)

ITERATION	OBJECTIVE	VAR(B)	VAR(A*B)	VAR(ERROR)
0	10.47318994	0.00000007	0.00000007	2.85000000
1	10.47318994	0.00000000	0.00000000	2.85000000

CONVERGENCE CRITERION MET

$\frac{1}{10}[\underline{y}'\underline{y} - R(\mu,\alpha)] = 28.5/10 = 2.85.$

P3V

DATA SET 1 - Random Model

```
** IN THIS VERSION OF BMDP3V -

-- THERE MUST BE AT LEAST ONE RANDOM COMPONENT
   IN ADDITION TO THE RESIDUAL.

-- IF REML (RESTRICTED MAXIMUM LIKELIHOOD) IS USED
   THE HYPOTHESIS OPTION IS NOT AVAILABLE.  MULTIPLE
   RUNS ARE REQUIRED TO TEST HYPOTHESES IN THIS CASE.
```

Control cards.

```
  PROGRAM CONTROL INFORMATION

/PROBLEM TITLE IS '2WAY X,RANDOM,DS1'.
/INPUT         VARIABLES ARE 3.
               FORMAT IS '(F2.0,F1.0,F2.0)'.
               CASES ARE 24.
/VARIABLE      NAMES ARE ROW,COLUMN,OBS.
/DESIGN        DEPENDENT IS OBS.
               RANDOM IS ROW.
               RANDOM IS COLUMN.
               RANDOM IS ROW,COLUMN.
               RNAMES ARE ROW,COLUMN,'ROW*COL'.
/END
```

Statement "METHOD = REML" is needed here when REML estimation is required.

Default method is ML.

Both ML and REML are iterative procedures. A maximum number of iterations can be provided for by the control statement MAXIT = n.

When running a series of data sets, the control statements for one set carry over to the next unless re-specified by new control statements.

```
  PROBLEM TITLE . . . . . .
2WAY X,RANDOM,DS1

  NUMBER OF VARIABLES TO READ IN. . . . . . . . . .        3
  NUMBER OF VARIABLES ADDED BY TRANSFORMATIONS. .          0
  TOTAL NUMBER OF VARIABLES . . . . . . . . . . .          3
  NUMBER OF CASES TO READ IN. . . . . . . . . . .         24
  CASE LABELING VARIABLES . . . . . . . . . . . .
  LIMITS AND MISSING VALUE CHECKED BEFORE TRANSFORMATIONS
  BLANKS ARE. . . . . . . . . . . . . . . . . . .      ZEROS
  INPUT UNIT NUMBER . . . . . . . . . . . . . . .          5
  REWIND INPUT UNIT PRIOR TO READING. . DATA. . .         NO
  NUMBER OF WORDS OF DYNAMIC STORAGE. . . . . . .      15000
  INPUT FORMAT. . . . .
(F2.0,F1.0,F2.0)


  VARIABLES TO BE USED
     1 ROW          2 COLUMN        3 OBS
```

VARIABLE NO.	NAME	BEFORE TRANSFORMATION MINIMUM LIMIT	MAXIMUM LIMIT	MISSING CODE	CATEGORY CODE	CATEGORY NAME	INTERVAL RANGE GREATER THAN	LESS THAN OR EQUAL TO
1	ROW				1.00000	* 1.0000		
					2.00000	* 2.0000		
					3.00000	* 3.0000		
2	COLUMN				1.00000	* 1.0000		
					2.00000	* 2.0000		
					3.00000	* 3.0000		
					4.00000	* 4.0000		

```
NOTE--CATEGORY NAMES BEGINNING WITH * WERE GENERATED BY THE PROGRAM.

  NUMBER OF CASES READ. . . . . . . . . . . . . .         24
```

P3V

DATA SET 1 - Random Model

DEPENDENT VARIABLE OBS

PARAMETER	ESTIMATE	STANDARD DEVIATION	EST/ST.DEV.	TWO-TAIL PROBABILITY (ASYMPTOTIC THEORY)
ERR.VAR.	20.0952380952	6.2015251128		
CONSTANT $\bar{y}_{\cdots}$ =	17.0000000000	2.4944382578	6.81516075	0.0
ROW	16.1547619048	15.2609703762		
COLUMN	0.0 $= \sigma^2_\beta$	0.0		
ROW*COL	0.0 $= \sigma^2_\gamma$	0.0		

-2*LOG(MAXIMUM LIKELIHOOD) 146.137726

VARIANCE-COVARIANCE MATRIX OF THE PARAMETERS

	ERR.VAR.	CONSTANT	ROW	COLUMN	ROW*COL
ERR.VAR.	38.4589				
CONSTANT	0.0	6.2222			
ROW	-4.8074	0.0	232.8972		
COLUMN	0.0	0.0	0.0	0.0	
ROW*COL	0.0	0.0	0.0	0.0	0.0

RESIDUAL ANALYSIS

CELL	OBSERVED OBS	PREDICTED OBS	ST.DEV. PRED.	OBSERVED-PREDICTED	ST.DEV. O-P	(O-P)/ ST.DEV.
	10.0000	17.0000	2.4944	-7.0000	5.4798	-1.277
	14.0000	17.0000	2.4944	-3.0000	5.4798	-0.547
	16.0000	17.0000	2.4944	-1.0000	5.4798	-0.182
	22.0000	17.0000	2.4944	5.0000	5.4798	0.912
	12.0000	17.0000	2.4944	-5.0000	5.4798	-0.912
	18.0000	17.0000	2.4944	1.0000	5.4798	0.182
	9.0000	17.0000	2.4944	-8.0000	5.4798	-1.460
	19.0000	17.0000	2.4944	2.0000	5.4798	0.365
	23.0000	17.0000	2.4944	6.0000	5.4798	1.095
	25.0000	17.0000	2.4944	8.0000	5.4798	1.460
	17.0000	17.0000	2.4944	0.0	5.4798	0.0
	21.0000	17.0000	2.4944	4.0000	5.4798	0.730
	24.0000	17.0000	2.4944	7.0000	5.4798	1.277
	32.0000	17.0000	2.4944	15.0000	5.4798	2.737
	18.0000	17.0000	2.4944	1.0000	5.4798	0.182
	24.0000	17.0000	2.4944	7.0000	5.4798	1.277
	13.0000	17.0000	2.4944	-4.0000	5.4798	-0.730
	17.0000	17.0000	2.4944	0.0	5.4798	0.0
	8.0000	17.0000	2.4944	-9.0000	5.4798	-1.642
	12.0000	17.0000	2.4944	-5.0000	5.4798	-0.912
	16.0000	17.0000	2.4944	-1.0000	5.4798	-0.182
	12.0000	17.0000	2.4944	-5.0000	5.4798	-0.912
	7.0000	17.0000	2.4944	-10.0000	5.4798	-1.825
	19.0000	17.0000	2.4944	2.0000	5.4798	0.365

$= \sqrt{6.222}$

ML estimation

With $\sigma^2_\beta = 0 = \sigma^2_\gamma$, the model reduces to the 1-way classification with 3 levels of α (see pages 9-10).

$\hat{\sigma}^2_e = 422/21 = 20.0952$

$\hat{\sigma}^2_\alpha = (448/3 - 20.0952)/8 = 16.1547.$

This is a', not a' - 1, because of ML. (See Corbeil and Searle, Biometrics, 779, 1976, Table 1.)

$\lambda = \log_e|\underset{\sim}{V}| + n + n\log_e 2\pi$ (see page 10)

$= \log_e[(\hat{\sigma}^2_e)^7(\hat{\sigma}^2_e + 8\hat{\sigma}^2_\alpha)]^3 + 24 + 24\log_e 2\pi$

$= 21\log_e(20.0952) + 3\log_e[20.0952 + 8(16.1547)]$

$+ 24 + 24(1.8378)$

$= 63.0101 + 15.018 + 68.1090$

$= 78.028 + 68.1090 = 146.137.$

$v(\hat{\sigma}^2_e) = 2\hat{\sigma}^4_e/a'(n'-1) = 2(20.0952)^2/21 = 38.4589.$

$v(\bar{y}_{\cdots}) = \hat{\sigma}^2_\alpha/3 + \hat{\sigma}^2_e/24$ (with $\hat{\sigma}^2_\beta = 0 = \hat{\sigma}^2_\gamma$)

$= 16.1547/3 + 20.0952/24 = 6.2222.$

$v(\hat{\sigma}^2_\alpha) = 2[(n'\hat{\sigma}^2_\alpha + \hat{\sigma}^2_e)^2/a' + \hat{\sigma}^4_e/a'(n'-1)]/n'^2$

$= \frac{2}{64}\left\{\frac{[8(16.1547) + 20.0952]^2}{3} + \frac{(20.0952)^2}{21}\right\}$

a', not a' - 1, as above.

$= 232.8972.$

$$\mathrm{cov}(\hat{\sigma}^2_\alpha, \hat{\sigma}^2_e) = \frac{-2\hat{\sigma}^4_e}{a'n'(n'-1)} = \frac{2(20.0952)^2}{3(8)7} = -4.807.$$

(see LM 416).

II

P3V

DATA SET 1 - Random Model

DEPENDENT VARIABLE OBS

PARAMETER	ESTIMATE	STANDARD DEVIATION	EST/ST.DEV.	TWO-TAIL PROBABILITY (ASYMPTOTIC THEORY)
ERR.VAR.	20.0952380952	6.2015251128		
CONSTANT	17.0000000000	3.0550504633	5.56455612	0.000
ROW	25.4880952381	26.0107286683		
COLUMN	0.0	0.0		
ROW*COL	0.0	0.0		

-2*LOG(MAXIMUM LIKELIHOOD) 139.104584

REML estimation

The data are balanced and so REML estimates and ANOVA estimates are the same.

$\hat{\sigma}^2_e$ = 20.0952 as in ML, see page II.

$\hat{\sigma}^2_\alpha = (448/2 - 20.0952)/8 = 25.4880.$

VARIANCE-COVARIANCE MATRIX OF THE PARAMETERS

	ERR.VAR.	CONSTANT	ROW	COLUMN	ROW*COL
ERR.VAR.	38.4589				
CONSTANT	0.0	9.3333			
ROW	-4.3074	0.0	784.6008		
COLUMN	0.0	0.0	0.0	0.0	
ROW*COL	0.0	0.0	0.0	0.0	0.0

$v(\hat{\sigma}^2_e)$ = 38.4589, as on page II.

$v(\bar{y}_{...}) = \hat{\sigma}^2_\alpha/3 + \hat{\sigma}^2_e/24$ (with $\hat{\sigma}^2_\beta = 0 = \hat{\sigma}^2_\gamma$)

$= 25.4880/3 + 20.0952/24 = 9.333.$

$v(\hat{\sigma}^2_\alpha) = 2[(n'\hat{\sigma}^2_\alpha + \hat{\sigma}^2_e)/a' + \hat{\sigma}^4_e/a'(n'-1)]/n'^2$

$$= \frac{2}{64}\left\{\frac{[8(25.4880) + 20.0592]^2}{2} + \frac{(20.0952)^2}{21}\right\}$$

$= 784.600$

$cov(\hat{\sigma}^2_\alpha, \hat{\sigma}^2_e) = -4.307$, as on page II.

SOURCE	SUM OF SQUARES	D.F.	MEAN SQUARE	F	PROBABILITY
CONSTANT	514.0203	1	514.0203	25.579	0.000
ERROR	381.8096	19	20.0952		

This is a pseudo analysis of variance table yielded by versions of BMDP3V dated prior to 1980. It is an attempted analysis of variance for the fixed effects in a mixed model (in this case, just μ), based on the asymptotic properties of the preceding variance-covariance matrix. BMD personnel advise that as of August 1, 1980, this table is being replaced by hypothesis tests for the fixed effects using the same variance-covariance matrix.

RESIDUAL ANALYSIS

CELL	OBSERVED OBS	PREDICTED OBS	ST.DEV. PRED.	OBSERVED-PREDICTED	ST.DEV. O-P	(O-P)/ ST.DEV.
	10.0000	17.0000	3.0551	-7.0000	6.0208	-1.163
	14.0000	17.0000	3.0551	-3.0000	6.0208	-0.498
	16.0000	17.0000	3.0551	-1.0000	6.0208	-0.166
	22.0000	17.0000	3.0551	5.0000	6.0208	0.830
	12.0000	17.0000	3.0551	-5.0000	6.0208	-0.830
	18.0000	17.0000	3.0551	1.0000	6.0208	0.166
	9.0000	17.0000	3.0551	-8.0000	6.0208	-1.329
	19.0000	17.0000	3.0551	2.0000	6.0208	0.332
	23.0000	17.0000	3.0551	6.0000	6.0208	0.997
	25.0000	17.0000	3.0551	8.0000	6.0208	1.329
	17.0000	17.0000	3.0551	0.0	6.0208	0.0
	21.0000	17.0000	3.0551	4.0000	6.0208	0.664
	24.0000	17.0000	3.0551	7.0000	6.0208	1.163
	32.0000	17.0000	3.0551	15.0000	6.0208	2.491
	18.0000	17.0000	3.0551	1.0000	6.0208	0.166
	24.0000	17.0000	3.0551	7.0000	6.0208	1.163
	13.0000	17.0000	3.0551	-4.0000	6.0208	-0.664
	17.0000	17.0000	3.0551	0.0	6.0208	0.0
	8.0000	17.0000	3.0551	-9.0000	6.0208	-1.495

P3V

DATA SET 1 - Mixed Model

(α's, rows, fixed)

CELL INFORMATION FOR VARIABLE CES

CELL	MEAN	ST.DEV.	COEF. OF VARIATION	COUNT	GROUPING VARIABLES ROW
1	15.00000	4.56696	0.30446	8.00000	* 1.0000
2	23.00000	4.65986	0.20260	8.00000	* 2.0000
3	13.00000	4.20283	0.32376	8.00000	* 3.0000

$\sqrt{\Sigma(x_i - \bar{x})^2/(n-1)}$ for each "cell", in this case, row.

ML estimation

With $\hat{\sigma}^2_B = 0 = \hat{\sigma}^2_Y$, the model reduces to $y_{ijk} = u + \alpha_i + e_{ijk}$.

DEPENDENT VARIABLE CES

PARAMETER	ESTIMATE		STANDARD DEVIATION	EST/ST.DEV.	TWO-TAIL PROBABILITY (ASYMPTOTIC THEORY)
ERR.VAR.	17.5833333333		5.0758711166		
CONSTANT	17.0000000000	$= \bar{y}_{...}$	0.2559432743	19.8611145	0.0
ROW	-2.0000000000	$\bar{y}_{i..} - \bar{y}_{...}$	1.2104865872	-1.65222740	0.098
ROW	6.0000000000		1.2104865872	4.95668411	0.000
COLUMN	0.0	$\hat{\sigma}^2_B = 0 = \hat{\sigma}^2_Y$	0.0		
ROW*COL	0.0		0.0		

$\hat{\sigma}^2_e = 422/24 = 17.5833$

ML estimation.

-2*LOG(MAXIMUM LIKELIHOOD) 136.915863

VARIANCE-COVARIANCE MATRIX OF THE PARAMETERS

	ERR.VAR.	CONSTANT	ROW	ROW	COLUMN	ROW*COL
ERR.VAR.	25.7645					
CONSTANT	0.0	0.7326				
ROW	0.0	0.0	1.4653			
ROW	0.0	0.0	-0.7326	1.4653		
COLUMN	0.0	0.0	0.0	0.0	0.0	
ROW*COL	0.0	0.0	0.0	0.0	0.0	0.0

$v(\hat{\sigma}^2_e) = 2(17.5833)^2/24 = 25.7645.$

$v(\bar{y}_{...}) = \hat{\sigma}^2_e/24 = 17.5833/24 = .7326.$

$v(\bar{y}_{i..} - \bar{y}_{...}) = \hat{\sigma}^2_e\left(\frac{1}{8} + \frac{1}{24} - \frac{1}{12}\right)$

$= 17.5833/12 = 1.4653.$

For $\delta_i = \bar{y}_{i..} - \bar{y}_{...}$

$\mathrm{cov}(\delta_i, \delta_{i'}) = \hat{\sigma}^2_e\left(\frac{1}{24} - \frac{1}{24} - \frac{1}{24}\right) = -.7326.$

CELL	GROUPING VARIABLES ROW	DUMMY VARIABLES		
1	* 1.0000	1.	1.	0.
2	* 2.0000	1.	0.	1.
3	* 3.0000	1.	-1.	-1.

CELL	OBSERVED MEAN	PREDICTED MEAN	SD.DEV. PRED.
1	15.0000	15.0000	1.4825
2	23.0000	23.0000	1.4825
3	13.0000	13.0000	1.4825

$= \sqrt{\hat{\sigma}^2_e/8} = \sqrt{17.5833/8}$

VARIANCE-COVARIANCE MATRIX OF PREDICTED CELL MEANS

	1	2	3
1	2.1979		

25

IV

P3V

DATA SET 2NB - Random Model

DEPENDENT VARIABLE Y

PARAMETER	ESTIMATE	STANDARD DEVIATION	EST/ST.DEV.	TWO-TAIL PROBABILITY (ASYMPTOTIC THEORY)
ERR.VAR.	6.2500000000	3.1250000000		
CONSTANT	16.0000000000	6.3639610307	2.51415730	0.012
ROW	77.0000000000	81.0987052916		
COL(ROW)	5.9166666667	8.0675318062		

ML estimation

See page 14.

-2*LOG(MAXIMUM LIKELIHOOD) 67.4436951

VARIANCE-COVARIANCE MATRIX OF THE PARAMETERS

	ERR.VAR.	CONSTANT	ROW	COL(ROW)
ERR.VAR.	9.7656			
CONSTANT	0.0	40.5000		
ROW	-0.0000	0.0	6577.0000	
COL(ROW)	-3.2552	0.0	-32.0000	65.0851

DEPENDENT VARIABLE Y

PARAMETER	ESTIMATE	STANDARD DEVIATION	EST/ST.DEV.	TWO-TAIL PROBABILITY (ASYMPTOTIC THEORY)
ERR.VAR.	6.2500000000	3.1250000000		
CONSTANT	16.0000000000	9.0000000000	1.77777767	0.075
ROW	157.9999999999	229.1375132974		
COL(ROW)	5.9166666667	8.0675318062		

-2*LOG(MAXIMUM LIKELIHOOD) 59.1127625

VARIANCE-COVARIANCE MATRIX OF THE PARAMETERS

	ERR.VAR.	CONSTANT	ROW	COL(ROW)
ERR.VAR.	9.7656			
CONSTANT	0.0	81.0000		
ROW	-0.0000	0.0	52504.0000	
COL(ROW)	-3.2552	0.0	-32.0000	65.0851

REML estimation

Same as ML except that

$$\hat{\sigma}^2_\alpha = \frac{1}{bn}\left[\frac{SSA}{a-1} + \frac{SSB(A)}{a(b-1)}\right]$$

a-1 for REML

a for ML

(See Corbeil and Searle, Biometrics, 779, 1976, Table 1.)

$$\hat{\sigma}^2_\alpha = \frac{1}{8}\left[\frac{972}{1} - \frac{48}{2}\right] = 158$$

= ANOVA estimate, see page 13.

P3V

DATA SET 2NU - Random Model

DEPENDENT VARIABLE Y

PARAMETER	ESTIMATE	STANDARD DEVIATION	EST/ST.DEV.	TWO-TAIL PROBABILITY (ASYMPTOTIC THEORY)
ERR.VAR.	3.5718975706	46.2960662040		
CONSTANT	13.7261878390	8.9403134220	1.53531361	0.125
ROW	147.7701904137	28.3373389266		
COL(ROW)	27.1201575637	35.5946112616		

-2*LOG(MAXIMUM LIKELIHOOD) 81.6737366

REML estimation

(i) REML(0), zero iterations (MAXIT = 0). Starting values for the iterative procedure are 0 for all σ^2's except σ_e^2. Therefore the resulting solutions to the REML equations for $\hat{\sigma}^2$ are the same as MINQUE0 estimates; i.e., these are MINQUE0 estimates (see p. 19

VARIANCE-COVARIANCE MATRIX OF THE PARAMETERS

	ERR.VAR.	CONSTANT	ROW	COL(ROW)
ERR.VAR.	2143.3257			
CONSTANT	0.0	79.9292		
ROW	9.8984	0.0	803.0046	
COL(ROW)	-823.5437	0.0	-479.0808	1266.9905

PARAMETER	ESTIMATE	STANDARD DEVIATION	EST/ST.DEV.	TWO-TAIL PROBABILITY (ASYMPTOTIC THEORY)
ERR.VAR.	20.0677758444	24.7190214198		
CONSTANT	13.9211208045	8.6915001292	1.60169315	0.109
ROW	146.1346408266	126.5560103271		
COL(ROW)	2.4825442857	30.6992097229		

-2*LOG(MAXIMUM LIKELIHOOD) 71.9684753

(i_a) REML(1), one iteration (MAXIT = 1).

These values are not MINQUE0 estimates — and are shown solely to emphasize that it is the zero'th iteration that yields MINQUE0 values.

PARAMETER	ESTIMATE	STANDARD DEVIATION	EST/ST.DEV.	TWO-TAIL PROBABILITY (ASYMPTOTIC THEORY)
ERR.VAR.	20.4824872861	10.3479597266		
CONSTANT	13.9294484257	8.9125714712	1.56289864	0.118
ROW	154.4728443447	226.0156144613		
COL(ROW)	1.1527117614	8.5933600948		

-2*LOG(MAXIMUM LIKELIHOOD) 68.4981689

(ii) After convergence

These are the REML estimates.

They are not ANOVA estimates because the data are unbalanced.

VARIANCE-COVARIANCE MATRIX OF THE PARAMETERS

	ERR.VAR.	CONSTANT	ROW	COL(ROW)
ERR.VAR.	107.0803			
CONSTANT	0.0	79.4339		
ROW	99.1790	0.0	51083.0547	
COL(ROW)	-36.5712	0.0	-209.9169	73.8458

27

P3V

This is a composite page.

DATA SET 2 - Random Model

DEPENDENT VARIABLE OBS

REML estimation

PARAMETER	ESTIMATE	STANDARD DEVIATION	EST/ST.DEV.	TWO-TAIL PROBABILITY (ASYMPTOTIC THEORY)
ERR.VAR.	29.0496220641	35.1865389945		
CONSTANT	10.6302524245	3.6651631141	2.90034866	0.004
ROW	15.5533807829	22.0771858540		
COLUMN	17.1963985765	19.5318883346		

-2*LOG(MAXIMUM LIKELIHOOD) 54.6726074

(i) Zero iteration

These are MINQUEO estimates (see pages 30-31).

VARIANCE-COVARIANCE MATRIX OF THE PARAMETERS

	ERR.VAR.	CONSTANT	ROW	COLUMN
ERR.VAR.	1238.0923			
CONSTANT	0.0	13.4334		
ROW	-483.7578	0.0	487.4021	
COLUMN	-377.1672	0.0	126.1779	381.4946

PARAMETER	ESTIMATE	STANDARD DEVIATION	EST/ST.DEV.	TWO-TAIL PROBABILITY (ASYMPTOTIC THEORY)
ERR.VAR.	4.1510519071	3.5200039367		
CONSTANT	8.7642378729	5.3162403272	1.64857769	0.099
ROW	46.3319960029	40.3247979160		
COLUMN	48.0189913108	49.8772382525		

(ii) After convergence

These are the REML estimates.

-2*LOG(MAXIMUM LIKELIHOOD) 50.3241862

VARIANCE-COVARIANCE MATRIX OF THE PARAMETERS

	ERR.VAR.	CONSTANT	ROW	COLUMN
ERR.VAR.	12.3504			
CONSTANT	0.0	28.2624		
ROW	-13.6853	0.0	1626.0891	
COLUMN	-11.8302	0.0	84.2421	2487.7388

SOURCE	SUM OF SQUARES	D.F.	MEAN SQUARE	F	PROBABILITY
CONSTANT	7.0511	1	7.0511	1.699	0.249
ERROR	20.7553	5	4.1511		

Ignore. See page III.

RESIDUAL ANALYSIS

CELL	OBSERVED OBS	PREDICTED OBS	ST.DEV. PRED.	OBSERVED-PREDICTED	ST.DEV. O-P	(O-P)/ ST.DEV.
	18.0000	8.7642	5.3162	9.2358	8.3809	1.102
	12.0000	8.7642	5.3162	3.2358	8.3809	0.386
	24.0000	8.7642	5.3162	15.2358	8.3809	1.818
	9.0000	8.7642	5.3162	0.2358	8.3809	0.028
	3.0000	8.7642	5.3162	-5.7642	8.3809	-0.688
	15.0000	8.7642	5.3162	6.2358	8.3809	0.744
	6.0000	8.7642	5.3162	-2.7642	8.3809	-0.330
	3.0000	8.7642	5.3162	-5.7642	8.3809	-0.688
	18.0000	8.7642	5.3162	9.2358	8.3809	1.102

HYPOTHESIS OPTION NOT AVAILABLE IF REML IS USED.

28

DATA SET 2 - Mixed Model
(a's, rows, fixed)

CELL INFORMATION FOR VARIABLE OBS

CELL	MEAN	ST.DEV.	COEF. OF VARIATION	COUNT	GROUPING VARIABLES ROW
1	18.00000	6.00000	0.33333	3.00000	* 1.0000
2	9.00000	0.0	0.0	1.00000	* 2.0000
3	9.00000	8.48528	0.94281	2.00000	* 3.0000
4	9.00000	7.93725	0.88192	3.00000	* 4.0000

DEPENDENT VARIABLE OBS

ML estimation

PARAMETER	ESTIMATE	STANDARD DEVIATION	EST/ST.DEV.	TWO-TAIL PROBABILITY (ASYMPTOTIC THEORY)
ERR.VAR.	2.0034145689	1.1566614257		
CONSTANT	8.5675715749	3.4012450441	2.51895142	0.012
ROW	9.4324284251	0.8162076577	11.5564070	0.0
ROW	-7.3838626790	1.2238087167	-6.03351021	0.0
ROW	-2.4809941712	0.9127664023	-2.71809769	0.007
COLUMN	33.7065260223	28.1151387719		

-2*LOG(MAXIMUM LIKELIHOOD) 43.5040438

VARIANCE-COVARIANCE MATRIX OF THE PARAMETERS

	ERR.VAR.	CONSTANT	ROW	ROW	ROW	COLUMN
ERR.VAR.	1.3379					
CONSTANT	0.0	11.5685				
ROW	0.0	-0.1653	0.6662			
ROW	0.0	0.3315	-0.4984	1.4977		
ROW	0.0	-0.0008	-0.1662	-0.5008	0.8331	
COLUMN	-0.4814	0.0	0.0	0.0	0.0	790.4605

CELL	GROUPING VARIABLES ROW	DUMMY VARIABLES			
1	* 1.0000	1.	1.	0.	0.
2	* 2.0000	1.	0.	1.	0.
3	* 3.0000	1.	0.	0.	1.
4	* 4.0000	1.	-1.	-1.	-1.

CELL	OBSERVED MEAN	PREDICTED MEAN	SD.DEV. PRED.
1	18.0000	18.0000	3.4502
2	9.0000	1.1837	3.7053
3	9.0000	6.0866	3.5214
4	9.0000	9.0000	3.4502

 ANOVA estimation from balanced data.

This is a composite page.

DATA SET 1

ANALYSIS OF VARIANCE FOR DEPENDENT VARIABLE 1

Random Model

	SOURCE	ERROR TERM	SUM OF SQUARES	D.F.	MEAN SQUARE	F	PROB.	EXPECTED MEAN SQUARE
1	MEAN		6936.000	1	6936.000			24(1) + 8(2) + 6(3) + 2(4) + (5)
2	ROW	RC	448.000	2	224.000	9.88	0.0126	8(2) + 2(4) + (5)
3	COL	RC	36.000	3	12.000	0.53	0.6784	6(3) + 2(4) + (5)
4	RC	I(RC)	136.000	6	22.667	1.09	0.4223	2(4) + (5)
5	I(RC)		250.000	12	20.833			(5)

ESTIMATES OF VARIANCE COMPONENTS

(1) 280.11111 = $\widehat{(\mu^2)} \neq (\hat{\mu})^2$
(2) 25.16667
(3) -1.77778
(4) 0.91667
(5) 20.83333

$(1) = (\mu^2)$
$(2) = \sigma^2_\alpha$
$(3) = \sigma^2_\beta$
$(4) = \sigma^2_\gamma$
$(5) = \sigma^2_e$

ANALYSIS OF VARIANCE FOR DEPENDENT VARIABLE 1

Mixed Model

	SOURCE	ERROR TERM	SUM OF SQUARES	D.F.	MEAN SQUARE	F	PROB.	EXPECTED MEAN SQUARE
1	MEAN	COL	6936.000	1	6936.000	578.00	0.0002	24(1) + 6(3) + (5)
2	ROW	RC	448.000	2	224.000	9.88	0.0126	8(2) + 2(4) + (5)
3	COL	I(RC)	36.000	3	12.000	0.58	0.6417	6(3) + (5)
4	RC	I(RC)	136.000	6	22.667	1.09	0.4223	2(4) + (5)
5	I(RC)		250.000	12	20.833			(5)

ESTIMATES OF VARIANCE COMPONENTS

(1) 288.50000
(2) 25.16667
(3) -1.47222
(4) 0.91667
(5) 20.83333

- -

GRAND MEAN 17.00000

CELL AND MARGINAL MEANS

R =	1	2	3
	15.00000	23.00000	13.00000

C =	1	2	3	4
	17.00000	16.00000	19.00000	16.00000

	C =	1	2	3	4
R =	1	12.00000	19.00000	15.00000	14.00000
	2	24.00000	19.00000	28.00000	21.00000
	3	15.00000	10.00000	14.00000	13.00000

Notes:

a. The symbol (2) here represents a quadratic form in the α_i's, the fixed effects. See Table 9.10, LM 403.

b. The model being used is the one that assumes the interactions add to zero (Σ-restrictions) over the levels of the fixed effect:

$$\gamma''_{\cdot j} = \sum_{i=1}^{a} \gamma''_{ij} = 0 ,$$

as in Table 9.10, LM 403. But the estimated interaction variance component is of the γ_{ij}'s and not of the γ''_{ij}'s, i.e.,

$$0.91667 = \hat{\sigma}^2_\gamma = \frac{a}{a-1}\hat{\sigma}^2_{\gamma''}$$

based on (38), LM 404.

c. Using the Σ-restrictions for the γ's, as shown in b, means that 2(4) is omitted from E(MSB) for columns – see Table 9.10, LM 403. Hence the estimate for $\hat{\sigma}^2_\beta$ is different from that in the random model:

Random: $\hat{\sigma}^2_\gamma = (12 - 22\tfrac{2}{3})/6 = -1.77777$

Mixed ($\gamma''_{\cdot j} = 0$): $\hat{\sigma}^2_\gamma = (12 - 20\tfrac{5}{6})/6 = -1.47222$.

This is a composite page.

DATA SET 1 - Random Model

DEPENDENT VARIABLE: Y

SOURCE		TYPE I EXPECTED MEAN SQUARE
ROW	E(MSA) =	VAR(ERROR) + 2 VAR(ROW*COL) + 8 VAR(ROW)
COL	E(MSB) =	VAR(ERROR) + 2 VAR(ROW*COL) + 6 VAR(COL)
ROW*COL	E(MSAB) =	VAR(ERROR) + 2 VAR(ROW*COL)

The data are balanced, and so mean squares Types I, II, III and IV are all the same.

SOURCE	TYPE II EXPECTED MEAN SQUARE
ROW	VAR(ERROR) + 2 VAR(ROW*COL) + 8 VAR(ROW)
COL	VAR(ERROR) + 2 VAR(ROW*COL) + 6 VAR(COL)
ROW*COL	VAR(ERROR) + 2 VAR(ROW*COL)

These are the expected values of these mean squares (ACO σ^2: SAS VARCOMP, p. 2).

SOURCE	TYPE III EXPECTED MEAN SQUARE
ROW	VAR(ERROR) + 2 VAR(ROW*COL) + 8 VAR(ROW)
COL	VAR(ERROR) + 2 VAR(ROW*COL) + 6 VAR(COL)
ROW*COL	VAR(ERROR) + 2 VAR(ROW*COL)

SOURCE	TYPE IV EXPECTED MEAN SQUARE
ROW	VAR(ERROR) + 2 VAR(ROW*COL) + 8 VAR(ROW)
COL	VAR(ERROR) + 2 VAR(ROW*COL) + 6 VAR(COL)
ROW*COL	VAR(ERROR) + 2 VAR(ROW*COL)

31

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	11	620.00000000	56.36363636	2.71	0.0507	0.712644	26.8491
ERROR	12	250.00000000	20.83333333		STD DEV		Y MEAN
CORRECTED TOTAL	23	870.00000000			4.56435465		17.00000000

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE II SS	F VALUE	PR > F
ROW	2	448.00000000	10.75	0.0021	2	448.00000000	10.75	0.0021
COL	3	36.00000000	0.58	0.6417	3	36.00000000	0.58	0.6417
ROW*COL	6	136.00000000	1.09	0.4223	6	136.00000000	1.09	0.4223

SOURCE	DF	TYPE III SS	F VALUE	PR > F	DF	TYPE IV SS	F VALUE	PR > F
ROW	2	448.00000000	10.75	0.0021	2	448.00000000	10.75	0.0021
COL	3	36.00000000	0.58	0.6417	3	36.00000000	0.58	0.6417
ROW*COL	6	136.00000000	1.09	0.4223	6	136.00000000	1.09	0.4223

T

DS1 MIXED MODEL, ROWS FIXED

GENERAL LINEAR MODELS PROCEDURE

DATA SET 1 - Mixed Model, Rows Fixed

QUADRATIC FORMS OF FIXED EFFECTS
IN THE EXPECTED MEAN SQUARES

DEPENDENT VARIABLE: Y

SOURCE: TYPE I MEAN SQUARE FOR ROW

	ROW 1	ROW 2	ROW 3
ROW 1	5.33333333	-2.66666667	-2.66666667
ROW 2	-2.66666667	5.33333333	-2.66666667
ROW 3	-2.66666667	-2.66666667	5.33333333

SOURCE: TYPE II MEAN SQUARE FOR ROW

	ROW 1	ROW 2	ROW 3
ROW 1	5.33333333	-2.66666667	-2.66666667
ROW 2	-2.66666667	5.33333333	-2.66666667
ROW 3	-2.66666667	-2.66666667	5.33333333

SOURCE: TYPE III MEAN SQUARE FOR ROW

	ROW 1	ROW 2	ROW 3
ROW 1	5.33333333	-2.66666667	-2.66666667
ROW 2	-2.66666667	5.33333333	-2.66666667
ROW 3	-2.66666667	-2.66666667	5.33333333

SOURCE: TYPE IV MEAN SQUARE FOR ROW

	ROW 1	ROW 2	ROW 3
ROW 1	5.33333333	-2.66666667	-2.66666667
ROW 2	-2.66666667	5.33333333	-2.66666667
ROW 3	-2.66666667	-2.66666667	5.33333333

SOURCE		TYPE I EXPECTED MEAN SQUARE
ROW	E(MSA)	= VAR(ERROR) + 2 VAR(ROW*COL) + Q(ROW)
COL	E(MSB)	= VAR(ERROR) + 2 VAR(ROW*COL) + 6 VAR(COL)
ROW*COL	E(MSAB)	= VAR(ERROR) + 2 VAR(ROW*COL)

Replaces $8\sigma_\alpha^2$

SOURCE	TYPE II EXPECTED MEAN SQUARE
ROW	VAR(ERROR) + 2 VAR(ROW*COL) + Q(ROW)
COL	VAR(ERROR) + 2 VAR(ROW*COL) + 6 VAR(COL)
ROW*COL	VAR(ERROR) + 2 VAR(ROW*COL)

SOURCE	TYPE III EXPECTED MEAN SQUARE
ROW	VAR(ERROR) + 2 VAR(ROW*COL) + Q(ROW)
COL	VAR(ERROR) + 2 VAR(ROW*COL) + 6 VAR(COL)
ROW*COL	VAR(ERROR) + 2 VAR(ROW*COL)

SOURCE	TYPE IV EXPECTED MEAN SQUARE
ROW	VAR(ERROR) + 2 VAR(ROW*COL) + Q(ROW)
COL	VAR(ERROR) + 2 VAR(ROW*COL) + 6 VAR(COL)
ROW*COL	VAR(ERROR) + 2 VAR(ROW*COL)

The data are balanced and so all of Types I, II, III and IV mean squares are the same. Because rows are fixed effects, E MS(rows) contains, for this mixed model, in place of σ_α^2 in the random model (page I), a quadratic in the α_i's. As shown in Table 9.9, LM 401, this quadratic is

$$\frac{bn}{a-1}\sum_{i=1}^{a}(\alpha_i-\bar{\alpha})^2 = 8\sum_{i=1}^{3}(\alpha_i-\bar{\alpha})^2/2 \quad \text{for } \bar{\alpha}=\tfrac{1}{3}\sum_{i=1}^{3}\alpha_i .$$

For the expected sum of squares this is

This matrix is this output.

$$8\sum_{i=1}^{3}(\alpha_i-\bar{\alpha})^2 = 8[\alpha_1\ \alpha_2\ \alpha_3][\underset{\sim}{I}_3-\tfrac{1}{3}\underset{\sim}{J}_3]\begin{bmatrix}\alpha_1\\ \alpha_2\\ \alpha_3\end{bmatrix} = [\alpha_1\ \alpha_2\ \alpha_3]\begin{bmatrix}5\frac{1}{3} & -2\frac{2}{3} & -2\frac{2}{3}\\ -2\frac{2}{3} & 5\frac{1}{3} & -2\frac{2}{3}\\ -2\frac{2}{3} & -2\frac{2}{3} & 5\frac{1}{3}\end{bmatrix}\begin{bmatrix}\alpha_1\\ \alpha_2\\ \alpha_3\end{bmatrix}.$$

It needs to be divided by $a-1=2$ to be the Q(row) of E (mean square) here.

See also page XIV.

32

This is a compos[illegible] page.

DS2NU (NESTED UNBALANCED) RANDOM MODEL

GENERAL LINEAR MODELS PROCEDURE

DATA SET 2NU - Random Model

DEPENDENT VARIABLE: Y

SOURCE	TYPE I EXPECTED MEAN SQUARE
ROW	VAR(ERROR) + 2.66666667 VAR(COL(ROW)) + 5.33333333 VAR(ROW)
COL(ROW)	VAR(ERROR) + 2.16666667 VAR(COL(ROW))

SOURCE	TYPE II EXPECTED MEAN SQUARE
ROW	VAR(ERROR) + 2.66666667 VAR(COL(ROW)) + 5.33333333 VAR(ROW)
COL(ROW)	VAR(ERROR) + 2.16666667 VAR(COL(ROW))

See page 15 of ACO σ^2: SAS VARCOMP.

SOURCE	TYPE III EXPECTED MEAN SQUARE
ROW	VAR(ERROR) + 1.76470588 VAR(COL(ROW)) + 4.23529412 VAR(ROW)
COL(ROW)	VAR(ERROR) + 2.16666667 VAR(COL(ROW))

SOURCE	TYPE IV EXPECTED MEAN SQUARE
ROW	VAR(ERROR) + 1.76470588 VAR(COL(ROW)) + 4.23529412 VAR(ROW)
COL(ROW)	VAR(ERROR) + 2.16666667 VAR(COL(ROW))

Type III is for the Σ-restricted model, and Type IV is the same, because it is a nested model.

This expectation is shown in equation (39) on page 16. The sum of squares is calculated for the Σ-restricted model; but its expected value is taken under the unrestricted model (see pages 15-21).

DEPENDENT VARIABLE: Y

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	4	924.00000000	231.00000000	10.64	0.0042	0.858736	42.3624
ERROR	7	152.00000000	21.71428571		STD DEV		Y MEAN
CORRECTED TOTAL	11	1076.00000000			4.65985898		11.00000000

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE II SS	F VALUE	PR > F
ROW	1	864.00000000	39.79	0.0004	1	864.00000000	39.79	0.0004
COL(ROW)	3	60.00000000	0.92	0.4786	3	60.00000000	0.92	0.4786

SOURCE	DF	TYPE III SS	F VALUE	PR > F	DF	TYPE IV SS	F VALUE	PR > F
ROW	1	564.94117647	26.02	0.0014	1	564.94117647	26.02	0.0014
COL(ROW)	3	60.00000000	0.92	0.4786	3	60.00000000	0.92	0.4786

See equation (13), page 8.

V

DS4 WITH INTERACTION, RANDOM MODEL

GENERAL LINEAR MODELS PROCEDURE

DATA SET 4 - Random Model
(with interaction)

DEPENDENT VARIABLE: Y

SOURCE		TYPE I EXPECTED MEAN SQUARE
ROW	$E\ MS(\alpha\mid\mu)$	= VAR(ERROR) + 2.13333333 VAR(ROW*COL) + 0.133333333 VAR(COL) + 4.8 VAR(ROW)
COL	$E\ MS(\beta\mid\mu,\alpha)$	= VAR(ERROR) + 1.74242424 VAR(ROW*COL) + 2.83333333 VAR(COL)
ROW*COL	$E\ MS(\gamma\mid\mu,\alpha,\beta)$	= VAR(ERROR) + 2.18181818 VAR(ROW*COL)

Same as page XVII, ACO σ^2: SAS VARCOMP.

SOURCE		TYPE II EXPECTED MEAN SQUARE
ROW	$E\ MS(\alpha\mid\mu,\beta)$	= VAR(ERROR) + 2.21818182 VAR(ROW*COL) + 4.4 VAR(ROW)
COL	$E\ MS(\beta\mid\mu,\alpha)$	= VAR(ERROR) + 1.74242424 VAR(ROW*COL) + 2.83333333 VAR(COL)
ROW*COL	$E\ MS(\gamma\mid\mu,\alpha,\beta)$	= VAR(ERROR) + 2.18181818 VAR(ROW*COL)

SOURCE	TYPE III EXPECTED MEAN SQUARE
ROW	VAR(ERROR) + 2.18181818 VAR(ROW*COL) + 4.36363636 VAR(ROW) = $E\ MS^*(\dot{\alpha}\mid\dot{\mu},\dot{\beta},\dot{\gamma})_\Sigma$
COL	VAR(ERROR) + 1.71698113 VAR(ROW*COL) + 2.81132075 VAR(COL) = $E\ MS^*(\dot{\beta}\mid\dot{\mu},\dot{\alpha},\dot{\gamma})_\Sigma$
ROW*COL	VAR(ERROR) + 2.18181818 VAR(ROW*COL)

taking expectations under the unrestricted model.

SOURCE	TYPE IV EXPECTED MEAN SQUARE
ROW	VAR(ERROR) + 2.18181818 VAR(ROW*COL) + 4.36363636 VAR(ROW)
COL	VAR(ERROR) + 1.83333333 VAR(ROW*COL) + 1.83333333 VAR(COL)
ROW*COL	VAR(ERROR) + 2.18181818 VAR(ROW*COL)

These are expectations under the random model of mean squares used in the fixed model for testing

$$H: \alpha_1 - \alpha_2 + \tfrac{1}{2}(\gamma_{11} + \gamma_{12} - \gamma_{21} - \gamma_{22}) = 0$$

$$H: \begin{bmatrix} 1 & 0 & -1 & 1 & 0 & -1 \\ 0 & 1 & -1 & 0 & 1 & -1 \end{bmatrix} [\beta_1\ \beta_2\ \beta_3\ \gamma_{11}\ \gamma_{12}\ \gamma_{13}]' = 0\ .$$

See pages 42 and 43, ACO: SAS GLM.

This is a composite page.

DATA SET 1 RANDOM MODEL WITH INTERACTION

DATA SET 1 - Random Model (with interaction)

LEAST-SQUARES ANALYSIS OF VARIANCE

Y

SOURCE	D.F.	SUM OF SQUARES	MEAN SQUARES	F	PROB>F
TOTAL	24	7806.000000			
TOTAL REDUCTION	12	7556.000000	629.666667	30.224	
MU-YM ← See page III	1	6936.000000	6936.000000	332.928	0.0000
ROW for explanation.	2	448.000000	224.000000	10.752	0.0021
COL	3	36.000000	12.000000	0.576	0.6417
ROW X COL	6	136.000000	22.666667	1.088	0.4223
REMAINDER	12	250.000000	20.833333 ← $\hat{\sigma}^2_e = 20\frac{5}{6}$		

See page 2

K FOR RANDOM EFFECTS COMPONENT (ROW)= 8.0000 DEGREES OF FREEDOM = 2.

Coefficient of σ^2_α in E(MS rows).

SS, CP, MS, MCP, VARIANCE AND COVARIANCE COMPONENTS

JOB	ROW	COL	RHM	RHM	SS OR CP	MS OR COV	COMPONENTS
1	1	1	Y	Y	448.00000000	224.00000000	25.39583333 ← $\hat{\sigma}^2_\alpha = 25\frac{19}{48}$

K FOR RANDOM EFFECTS COMPONENT (COL)= 6.0000 DEGREES OF FREEDOM = 3.

SS, CP, MS, MCP, VARIANCE AND COVARIANCE COMPONENTS

JOB	ROW	COL	RHM	RHM	SS OR CP	MS OR COV	COMPONENTS
1	1	1	Y	Y	36.00000000	12.00000000	-1.47222222 ← $\hat{\sigma}^2_\beta = -1\frac{17}{36}$

K FOR RANDOM EFFECTS COMPONENT (ROW*COL)= 4.5714 DEGREES OF FREEDOM = 6.

SS, CP, MS, MCP, VARIANCE AND COVARIANCE COMPONENTS

JOB	ROW	COL	RHM	RHM	SS OR CP	MS OR COV	COMPONENTS
1	1	1	Y	Y	136.00000000	22.66666667	0.40104167 ← $\hat{\sigma}^2_\gamma$ is wrong.

NORMAL TERMINATION OF PROCEDURE HARVEY---PROBLEM NO. 1

This coefficient is not correct.

SAS HARVEY is not designed for handling interactions in mixed models. This has been confirmed (October 1, 1980) with the program developer, W. R. Harvey, Dairy Science Department, Ohio State University.

This is a composite page.

DATA SET 1 RANDOM MODEL NO INTERACTION

LEAST-SQUARES ANALYSIS OF VARIANCE

DATA SET 1 - Random Model (no interaction)

Y

SOURCE	D.F.	SUM OF SQUARES	MEAN SQUARES	F	PROB>F
TOTAL	24	7806.000000			
TOTAL REDUCTION	6	7420.000000	1236.666667	57.668	
MU-YM ← See page III for explanation.	1	6936.000000	6936.000000	323.440	0.0000
ROW	2	448.000000	224.000000	10.446	0.0010
COL	3	36.000000	12.000000	0.560	0.6485
REMAINDER	18	386.000000	21.444444 ← $\hat{\sigma}^2_e = 21\frac{4}{9}$		

See page 2 (bottom)

K FOR RANDOM EFFECTS COMPONENT (ROW)= 8.0000 DEGREES OF FREEDOM = 2.

SS, CP, MS, MCP, VARIANCE AND COVARIANCE COMPONENTS

JOE	ROW	COL	RHM	RHM	SS OR CP	MS OR COV	COMPONENTS
1	1	1	Y	Y	448.00000000	224.00000000	25.31944444 ← $\hat{\sigma}^2_\alpha = 25\frac{23}{72} = \hat{\sigma}^2_\alpha$

K FOR RANDOM EFFECTS COMPONENT (COL)= 6.0000 DEGREES OF FREEDOM = 3.

SS, CP, MS, MCP, VARIANCE AND COVARIANCE COMPONENTS

JOE	ROW	COL	RHM	RHM	SS OR CP	MS OR COV	COMPONENTS
1	1	1	Y	Y	36.00000000	12.00000000	-1.57407407 ← $\hat{\sigma}^2_\beta = -1\frac{31}{54} = \hat{\sigma}^2_\beta$

NORMAL TERMINATION OF PROCEDURE HARVEY---PROBLEM NO. 1

In the mixed model, with α's fixed, $\hat{\sigma}^2_e$ and $\hat{\sigma}^2_\beta$ are the same as here.

This is a comp[...]e page.

DATA SET 2NB(NESTED BALANCED) RANDOM MODEL
USER MEAN = 0

Columns within rows are treated individually, within each row.

LEAST-SQUARES ANALYSIS OF VARIANCE

DATA SET 2NB - Random Model

Y

SOURCE	D.F.		SUM OF SQUARES	MEAN SQUARES	F	PROB>F
TOTAL	12	SST	=4142.000000			
TOTAL REDUCTION	4	SST - SSE	=4092.000000	1023.000000	163.680	
MU-YM	1	$N\bar{y}^2$	=3072.000000	3072.000000	491.520	0.0000
ROW	1	SSA	= 972.000000	972.000000	155.520	0.0000
COL	1	SSB(A)	= 24.000000	24.000000	3.840	0.0857
COL	1		24.000000	24.000000	3.840	0.0857
REMAINDER	8	SSE	= 50.000000	6.250000		

Specification of YM

The user can specify an estimate of $\bar{y}_{...}$, denoted YM. If YM is not specified, it is taken as zero, in which case sums of squares are as shown.

Y

USER MEAN = 15

Output for YM = 15

SOURCE	D.F.	SUM OF SQUARES	MEAN SQUARES	F	PROB>F
TOTAL	12	1082.000000			
TOTAL REDUCTION	4	1032.000000	258.000000	41.280	
MU-YM	1	12.000000	12.000000	1.920	0.2033
ROW	1	972.000000	972.000000	155.520	0.0000
COL	1	24.000000	24.000000	3.840	0.0857
COL	1	24.000000	24.000000	3.840	0.0857
REMAINDER	8	50.000000	6.250000		

$\Sigma(y_{ijk} - YM)^2 = SST + N(YM)^2 - 2N(YM)\bar{y}_{..}$
$= 4142 + 12(15^2) - 2(12)(15)(16)$
$= 4142 + 2700 - 5760 = 1082.$

$\Sigma(y_{ijk} - YM)^2 - SSE = 1082 - 50 = 1032.$

$N(\bar{y} - YM)^2 = 12(16 - 15)^2 = 12.$

Y

USER MEAN = 16

SOURCE	D.F.	SUM OF SQUARES	MEAN SQUARES	F	PROB>F
TOTAL	12	1070.000000			
TOTAL REDUCTION	4	1020.000000	255.000000	40.800	
MU-YM	1	0.0	0.0	0.0	1.0000
ROW	1	972.000000	972.000000	155.520	0.0000
COL	1	24.000000	24.000000	3.840	0.0857
COL	1	24.000000	24.000000	3.840	0.0857
REMAINDER	8	50.000000	6.250000		

Output for YM = 16

K FOR RANDOM EFFECTS COMPONENT (ROW)= 6.0000 DEGREES OF FREEDOM = 1.

SS, CP, MS, MCP, VARIANCE AND COVARIANCE COMPONENTS

JOB	ROW	COL	RHM	RHM	SS OR CP	MS OR COV	COMPONENTS
20	1	1	Y	Y	972.00000000	972.00000000	$\hat{\sigma}^2_\alpha$ = 160.95833333 $= (972 - 6\frac{1}{4})/6 = 160\frac{23}{24}$.

K FOR RANDOM EFFECTS COMPONENT (COL)= 3.0000 DEGREES OF FREEDOM = 2.

See page 3

SS, CP, MS, MCP, VARIANCE AND COVARIANCE COMPONENTS

For the mixed model, with rows fixed, only this and $\hat{\sigma}^2_e = 6\frac{1}{4}$ are calculated.

JOB	ROW	COL	RHM	RHM	SS OR CP	MS OR COV	COMPONENTS
20	1	1	Y	Y	48.00000000	24.00000000	$\hat{\sigma}^2_\beta$ = 5.91666667 $= (24 - 6\frac{1}{4})/3 = 5\frac{11}{12}$.

This is a composite page.

DATA SET 2NU(NESTED UNBALANCED) RANDOM MODEL

LEAST-SQUARES ANALYSIS OF VARIANCE

DATA SET 2NU - Random Model

Y

SOURCE	D.F.	SUM OF SQUARES	MEAN SQUARES	F	PROB>F
TOTAL	12	2528.000000			
TOTAL REDUCTION	5	2376.000000	475.200000	21.884	
MU-YM	1	1620.941176	1620.941176	74.649	0.0001
ROW	1	564.941176	564.941176	26.017	0.0014
COL	1	12.000000	12.000000	0.553	0.4814
COL	2	48.000000	24.000000	1.105	0.3827
REMAINDER	7	152.000000	21.714286		

$60\{$ 12.000000 + 48.000000

$\hat{\sigma}^2_e = 21\frac{5}{7}$, as on page 6.

$R(\beta{:}\alpha|\mu,\alpha)$, as on page 6 .

$R^*(\dot{\alpha}|\dot{\mu},\dot{\beta}{:}\dot{\alpha})_\Sigma$, due to rows, in the Σ-restricted model. [See pages 7-10, especially equation (14) on page 10; and also page 11.]

$R^*(\mu|\alpha,\beta{:}\alpha)_\Sigma$ [See page 11.]

K FOR RANDOM EFFECTS COMPONENT (ROW)= 4.2353 DEGREES OF FREEDOM = 1.

$= 4\frac{4}{17}$ = coefficient of σ^2_α in E $R^*(\dot{\alpha}|\dot{\mu},\dot{\beta}{:}\dot{\alpha})_\Sigma$ under the restricted model. [See equation (28), page 16.]

SS, CP, MS, MCP, VARIANCE AND COVARIANCE COMPONENTS

JOB	ROW	COL	RHM	RHM	SS OR CP	MS OR COV	COMPONENTS
20	1	1	Y	Y	564.94117647	564.94117647	128.26190476

$R^*(\dot{\alpha}|\dot{\mu},\dot{\beta}{:}\dot{\alpha})_\Sigma$

$$\frac{R^*(\dot{\alpha}|\dot{\mu},\dot{\beta}{:}\dot{\alpha})_\Sigma/1 - 21\frac{5}{7}}{4\frac{4}{17}} =$$

$\hat{\sigma}^2_\alpha$ under the restricted model. [See equation (29), page 16.]

K FOR RANDOM EFFECTS COMPONENT (COL)= 2.1667 DEGREES OF FREEDOM = 3.

$= 2\frac{1}{6}$ = coefficient of σ^2_β in $R^*(\beta{:}\alpha|\mu,\alpha) = R(\beta{:}\alpha|\mu,\alpha)$.

SS, CP, MS, MCP, VARIANCE AND COVARIANCE COMPONENTS

JOB	ROW	COL	RHM	RHM	SS OR CP	MS OR COV	COMPONENTS
20	1	1	Y	Y	60.00000000	20.00000000	-0.79120879

NORMAL TERMINATION OF PROCEDURE HARVEY---PROBLEM NO. 20

$$\frac{20 - 21\frac{5}{7}}{2\frac{1}{6}} = -\frac{72}{91} = \hat{\sigma}^2_\beta$$, see page 6.

NEW YORK STATE COLLEGE OF AGRICULTURE AND LIFE SCIENCES
CORNELL UNIVERSITY

BIOMETRICS UNIT
IN THE
DEPARTMENT OF PLANT BREEDING AND BIOMETRY

Telephone — 607-256-5488

Mailing Address:
337 Warren Hall, Cornell University
Ithaca, New York 14853

ANNOTATED COMPUTER OUTPUT (ACO)

for ANALYSIS OF VARIANCE and for VARIANCE COMPONENTS ESTIMATION

An Annotated Computer Output (ACO) is a manual for helping statisticians understand the output from statistical computing routines, especially routines designed for analysis of variance calculations and for variance components estimation, particularly when those routines are used on unbalanced data (i.e., unequal-subclass-numbers data). The ACO for a computing routine consists of output from processing small sets of hypothetical, unbalanced, data having differing numbers of empty cells. The output has been annotated with extensive notes, comments and description, using notation and methodology of Linear Models, S. R. Searle, Wiley, 1971. Each ACO is authored by S. R. Searle and a colleague, is of some 30-70 pages in length, of $8\frac{1}{2}$ x 11 size, and includes the data sets and their basic analysis of variance calculations.

ACO's are available for analysis of variance calculations for fixed effects models, and for variance components estimation for mixed and random models. For the fixed effects models, two of the data sets include covariates. For the variance components calculations, three special data sets are included so that the data sets cover a range of balanced and unbalanced data, of crossed and nested classifications, of all cells filled and some empty, and of data with and without interaction. Hand calculations and computer output are given in this case both for random models and for mixed models, and many details of the derivation of individual output values are also shown.

An ACO is available for the following routines:

Fixed Effects Models

BMDP2V, 25 pages

GENSTAT (ANOVA and REGRESSION), 76 pages

MOPT (MINITAB, OSIRIS, P-STAT and TROLL) 34 pages

RUMMAGE, 29 pages

SAS GLM, 67 pages

SAS HARVEY, 59 pages

SPSS (ANOVA), 34 pages

Variance Components

BMDP-V (P3V, P2V and P8V), 60 pages

SAS GLM VARCOMP, 59 pages

SAS HARVEY, for σ^2 (approx. 40 pages)

SAS RANDOM (and NESTED), 47 pages

Upon receipt by the Biometrics Unit of $5.00 U.S. for each ACO ordered (cash, or a check made payable to Cornell University), orders will be promptly mailed, post-free.

ORDER FORM

Over.

40

ORDER FORM

To: Biometrics Unit
339 Warren Hall
Cornell University
Ithaca, New York 14853
U. S. A.

Please mail Annotated Computer Output (ACO) as follows:

For fixed effects (analysis of variance) models

_____ copies of BMDP2V

_____ copies of GENSTAT (ANOVA and REGRESSION)

_____ copies of MOPT (MINITAB, OSIRIS, P-STAT and TROLL)

_____ copies of RUMMAGE

_____ copies of SAS GLM

_____ copies of SAS HARVEY

_____ copies of SPSS (ANOVA)

For variance components and mixed models

_____ copies of BMDP-V (P3V, P2V and P8V)

_____ copies of SAS GLM VARCOMP

_____ copies of SAS HARVEY, for σ^2

_____ copies of SAS RANDOM (and NESTED)

TOTAL: _____ copies

They are to be sent to:

____

____

______________________________ (please print)

____

____

Enclosed is $ __________ ($5.00 per copy).

NOTE: An order will be mailed only after the check for it has been received. This is the only invoice that will be used.